

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:45:08 ; Search time 25.29 Seconds
(without alignments)
890.486 Million cell updates/sec

Title: US-09-762-926-4
Perfect score: 4904
Sequence: 1 MRSSFLKPKICFLMGVLY.....SVLTNFARGRFLTIMSKYF 922

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0*
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	284	5.8	915	3	US-08-613-009A-18
2	278	5.7	915	1	US-08-487-890A-96
3	278	5.7	915	2	US-08-363-124A-2
4	278	5.7	915	2	US-08-478-435-96
5	278	5.7	915	2	US-08-337-483-96
6	278	5.7	915	2	US-08-478-373-96
7	278	5.7	915	3	US-08-474-671-96
8	278	5.7	915	3	US-08-483-577A-96
9	278	5.7	915	4	US-08-897-438-96
10	278	5.7	915	4	US-08-637-654-96
11	269.5	5.5	908	1	US-08-487-890A-94
12	269.5	5.5	908	2	US-08-478-435-94
13	269.5	5.5	908	2	US-08-337-483-94
14	269.5	5.5	908	2	US-08-478-373-94
15	269.5	5.5	908	3	US-08-474-671-94
16	269.5	5.5	908	3	US-08-483-577A-94
17	269.5	5.5	908	3	US-08-448-194-4
18	269.5	5.5	908	3	US-08-613-009A-16
19	269.5	5.5	908	4	US-08-897-438-94
20	269.5	5.5	908	4	US-08-867-921-4
21	269.5	5.5	908	4	US-08-637-654-94
22	261	5.3	909	2	US-08-363-124A-4
23	260	5.3	791	3	US-08-537-361E-4
24	257	5.2	791	4	US-08-817-707-4
25	257	5.2	911	1	US-08-487-890A-95
26	257	5.2	911	2	US-08-478-435-95
27	257	5.2	911	2	US-08-337-483-95

28	257	5.2	911	2	US-08-478-373-95	Sequence 95, Appl
29	257	5.2	911	3	US-08-474-671-95	Sequence 95, Appl
30	257	5.2	911	3	US-08-483-577A-95	Sequence 95, Appl
31	257	5.2	911	3	US-08-448-194-6	Sequence 6, Appl
32	257	5.2	911	3	US-08-613-009A-17	Sequence 17, Appl
33	257	5.2	911	4	US-08-897-438-95	Sequence 95, Appl
34	257	5.2	911	4	US-08-867-921-6	Sequence 6, Appl
35	257	5.2	911	4	US-08-637-654-95	Sequence 95, Appl
36	255	5.2	790	4	US-08-817-707-6	Sequence 6, Appl
37	250	5.1	884	1	US-08-066-167-2	Sequence 2, Appl
38	243.5	5.0	941	4	US-09-074-658-75	Sequence 75, Appl
39	242	4.9	790	4	US-08-817-707-8	Sequence 8, Appl
40	239	4.9	790	3	US-08-537-361E-8	Sequence 15, Appl
41	239	4.9	912	3	US-08-613-009A-15	Sequence 15, Appl
42	238.5	4.9	792	3	US-08-537-361E-6	Sequence 6, Appl
43	237.5	4.8	792	1	US-08-326-670A-2	Sequence 2, Appl
44	237.5	4.8	792	3	US-08-990-470A-2	Sequence 2, Appl
45	237.5	4.8	792	4	US-08-817-707-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-613-009A-18
; Sequence 18, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,009A
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-542
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-613-009A-18

Query Match 5.8%; Score 284; DB 3; Length 915;
Best Local Similarity 21.5%; Pred. No. 1.3e-15;
Matches 225; Conservative 128; Mismatches 321; Indels 374; Gaps 59;

QY 5 FRLKPICFYLMGVMLYHHSYAEDAGRAGSEAOIQVLEDDVHVAK--RVPKDKKVFDTARA 62
Db 7 FRLNLCLSLMTAL---PAYAENV-QAG-QAQBKQLDIOVKAKKQKTRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIFC-AFTQODK--SSGIVSLNIRGDSGFRVNTMVDGIT 118
Db 62 VKTADTLKSEQVLDIRDLTRYDFPCIAVVEQGRASSG---YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TFYSTSTAG--RAGSSQFASVDSNFIAGLDVVVKGSFSGAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGTRTAGSGGAINIEYENVKAVEISKSGNSVEQSGALAGSVAFOTKT 177
QY 176 VDDVQGNNTYGLLLK-GLTGTSN--TKGNAMAA-IGARKWL--ESGASVGLYGHRSRS 229
Db 178 ADDVIGEGRWGQIQSKTAYSGKNRGLTQSLALAGRIGCAEALLIRTGRAHRAH--- 233
QY 230 VAQYRVGGGQHIGNFGAELYERKORYFVQEGGLFNFSNKWDRDFORPYKWKYQ 289
Db 234 -----EAGRGVQSFNPLAPVDDGSKYAFIVEECK-----NGHEK-----C 272
QY 290 KYNDPQELQKIEGHRK---SWRENLAPOVDITPIDPSSLKOOS-----AGNLF--KLEY 339
Db 273 KANPPKD---VVGEDKQVSTEDYTGPNRFLA--DPLSVESRSLFRPGFREKNRHY 326
QY 340 DGVENKTAQ---PRDLNLTIGSKII-----NRNQNFGYGLSLNSYANLNL 383
Db 327 IGGILERTQOTFTDRMTVPFAFLTKAVFDANQKQAGSLRGKYGAGNH-----KYGGL-- 379
QY 384 TAAYNSGRQYKPGSKFTGGLLKDFETYNNAKILDLNNTATFPLPRETELQTLGNYF 443
Db 380 ---FTSGENNAVGAEY-GGVFYD-----ETHKSYRGLEYV 413
QY 444 HNEYGKRRFPPEELGFPD--GPDODNGL-----YSYLCRFRKGDK 480
Db 414 YTNADKDTWADYARLSYDRQIGLDNFQOQTHCSADGSKYCRPSADKPSY---YKSDR 470
QY 481 GLLPKSTIVQPGSQVNTFYFPAALKKDIYRLNYSNTNFGY-RFCG-----EYTYGG 534
Db 471 VIYGESKLLQAAFKKSFDTAKI-----RHNLSVN-LGYDRFGSLNRHQDY---YYQ 518
QY 535 SDDEFKRAFGENSPYKHKCNOSGCIYEPVLKYGKKRA-----NHNYSVISAD-- 583
Db 519 SAN---RAYSKTTP-----QNGKKTSPNGRKKNPYVWSIGRGNVTR 559
QY 584 ---FGD-----YFMPFA-----YSRTH----- 598
Db 560 QICLFGNNTVDCPRINGSKYAAYVRNVLGRWADVAGLRYDYRSTHSDDGSVSTG 619
QY 599 -----RMPNIQEMFYQIGDSGVHTALKPERANTWQ 629
Db 620 THRTLWSNAGIVLPADWLDLTYRTSTGFRPLPSFAEMYGWRSGDKIKAVKIDPEKSFNE 679
QY 630 FG-----FNTYKGLKODDITLGLKLVGYRSRIDN-----YIH--- 662
Db 680 AGIVFKGDFGNLEASWFNNAYRDLIV-----GYEAOIKDGKEQVKGNGPAYLNAQS 730
QY 663 ---NVYK--W---WDLGNIP-SWVSSTGLAYT-IQHRNFKDKVHKHGFELNLYD 709
Db 731 ARITGINILCKIDWNGYD--KLPEGWYST--PAYNRVRVRIKIKRDRDITQSHL-FD 784
QY 710 YGFEFTNLAYOKSTOPTNFSASPNASKEDQLKOGYGLSRVSALPRDYGRLEVGT 769
Db 785 -----AIQPSRYV-----VGSYDQPEGKGVNGMLT-----YSKAKEIT 819
QY 770 RWLGNKLTLCGAMRYFCKSRATAERYIDGTNGGNTSNVRLQCKRSIKOTETLARQPLI 829
Db 820 ELGSRALLNG-----NSRNTKATSRRTRPWYI 847
QY 830 FDEYAAEYEPKKNLIFRAEVENKLFDRRYI 857
Db 848 VDVSGYTYVKKHFTLRAGVYNLLNHRIV 875

RESULT 2
US-08-487-890A-96
; Sequence 96, Application US/08487890A
; Patent No. 5708149
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,890A
; FILING DATE: 07-JUN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-487-890A-96

Query Match 5.7%; Score 278; DB 1; Length 915;

Best Local Similarity 21.4%; Pred. No. 4.4e-15;

Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

QY 5 FRLKPICFYLMGVMLYHHSYAEDAGRAGSEAOIQVLEDDVHVAK--RVPKDKKVFDTARA 62
Db 7 FRLNLCLSLMTAL---PAYAENV-QAG-QAQBKQLDIOVKAKKQKTRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIFC-AFTQODK--SSGIVSLNIRGDSGFRVNTMVDGIT 118
Db 62 VKTADTLKSEQVLDIRDLTRYDFPCIAVVEQGRASSG---YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TFYSTSTAG--RAGSSQFASVDSNFIAGLDVVVKGSFSGAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGTRTAGSGGAINIEYENVKAVEISKSGNSVEQSGALAGSVAFOTKT 177
QY 176 VDDVQGNNTYGLLLK-GLTGTSN--TKGNAMAA-IGARKWL--ESGASVGLYGHRSRS 229
Db 178 ADDVIGEGRWGQIQSKTAYSGKNRGLTQSLALAGRIGCAEALLIRTGRAHRAH--- 233

```
QY 230 VAQNYRVGGGQHQHIGNFGAEYLERRKORYFVQEGGLKFNSNGKWERDQFORPYWKTWYQ 289
Db 234 -----EAGRGVGSFNLAPVDGSKYAFIVEECK-----NGHEKCKAMP-----276
QY 290 KYNDDPQELQYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KDDVGEDKQVSTFDTYGNRFLA--DPLSYESRSWLFRCGFRFENKRYH 326
QY 340 DGVFNKYTAQ---FRDLNTKIGSRKII-----NRNYQFNYGLSLNSYANLNL 383
Db 327 IGGILERTQOTFTDTRDMTVPFAFLTKAQVFNKQAGSLRNGKYAGNH-----KYGGL--379
QY 384 TAAVNSGRQYKPGSKFTGCLLKDFETYNNAKILDLNNTATFRLPRETELQTLGNYF 443
Db 380 ---FTSGENNAVGAEY-GTGVFYD-----ETHTKRYGLEVY 413
QY 444 HNEYGKNRFPPEELGLFFD--GPDQDNGL-----YSYLGRKFGDK 480
Db 414 YTNADKDTWADYARLSYDRQIGLDNHFQTHCSADGSKYCRPSADKPSY---YASDR 470
QY 481 GLLPQKSTIVOPAGSVFNTFYFDAALKKDIYRLNYSNTVGY-RFGG-----EYTGYYG 534
Db 471 VIYGESHKLLQAAFKKSFTAKI-----RHLNSVN-LGYDRFGSGLNRHODY--YYQ 518
QY 535 SDDEFKAFGENSPYTKKHCHNCQSGIYEPVLKKGKRA-----NNHSYSISAD-----583
Db 519 SAN---RAYSLKTPP-----QNNGKTPNGREKNPYWISIGRGNVYTR 559
QY 584 ---FGD-----YFMPPA-----SYSRTH-----598
Db 560 QICLFGNNTYDCTPRINSKSYAAVRDNVRLGRWADVGAGLRYDYRSTHSDDGSVTG 619
QY 599 -----RMPNIQEMYFSQIGDGVHTALKPERANTWQ 629
Db 620 THRTLSNAGIVLKPADWLDITYRSTGFRPLSPFAEMYGRSGDKIKAVKIDPEKSFKE 679
QY 630 FG-----FNTYKGLKQDDTLGLKLYCYRSRDN-----YIH---662
Db 680 AGIVKGDGPNLEASFENNAYRDLIV-----GYEAQIKDGKEQVKGNGPAYLNAQS 730
QY 663 -----NVYKG--W---WDLGNIP-SWVSSTGLAYT-IQHRNFKDKVHKHGFLELNYD 709
Db 731 ARITGINILKIDWGVMD---KLPEGWYST--PAYNRVVRDIKKRADRTDIQSHL-FD 784
QY 710 YGRFTNLSTAYOKSTOPTWFSASEPNASKEDQLKQYGLSRVSALPRDYGRLEVGT 769
Db 785 -----AIQPSRYV-----VCSGYDQPEGKGVNGMLT-----YSKAKEIT 819
QY 770 RWLGNKLTGAMRYFGKSIIRATAEERYIDGTNGGNTSNVRLQKRSIKQTEILARQPLI 829
Db 820 ELGSRALLNG-----NSRNTKATARTRPWYI 847
QY 830 FDFVAAEYPRKKNLIFRAEVKNLFDRRYI 857
Db 848 VDVSYYTVKKHFTLRAGVYNLLNHRVY 875

RESULT 3
US-08-363-124A-2
; Sequence 2, Application US/08363124A
; Patent No. 5912336
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
```

```
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,124A
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,254
; FILING DATE: 20-SEP-1993
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-363-124A-2
```

Query Match 5.7% Score 278 DB 2 Length 915;

Best Local Similarity 21.4%; Pred. No. 4.4e-15;

Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

```
QY 5 FRLPICPYLMGVNLYHSHYAEDAGRASEAQIOVLEDVHVYKAK--RVPDKKVFETDARA 62
Db 7 FRNLILSLMTAL---PAYAENV-QAG-QAQEKQDLOTVQKAKKQKTRDRNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSFGFVNTWVDGIT 118
Db 62 VKTADTLSEQVLDIRLDTRYDPIGVAVVEQGRASSG---YSIRG-MDKNRVSTVDGLA 117
QY 119 Q-TFYSTSDAG--RAGSSQFSGASVDSNFTAGLDVVKGSPGSGAGINSLAGSANLRTLG 175
Db 118 QIOSYTAQAALGGTRTAGSSGAINETENYENKVAEISKGSNSVEQSGGALAGSAFQTKT 177
QY 176 VDDVVOGNNTYGLLLK-GLTGTNS--TKGNAMAA-IGARKWL--ESGASGVLYGHRSRS 229
Db 178 ADDVIGEGRWGIQSKTAYSGKNRGLTQSIALAGRIGGAELLRTGRHAGEIRAH----233
QY 230 VAQNYRVGGGQHQHIGNFGAEYLERRKORYFVQEGGLKFNSNGKWERDQFORPYWKTWYQ 289
Db 234 -----EAGRGVGSFNLAPVDGSKYAFIVEECK-----NGHEKCKAMP-----276
QY 290 KYNDDPQELQYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KDDVGEDKQVSTFDTYGNRFLA--DPLSYESRSWLFRCGFRFENKRYH 326
QY 340 DGVFNKYTAQ---FRDLNTKIGSRKII-----NRNYQFNYGLSLNSYANLNL 383
Db 327 IGGILERTQOTFTDTRDMTVPFAFLTKAQVFNKQAGSLRNGKYAGNH-----KYGGL--379
QY 384 TAAVNSGRQYKPGSKFTGCLLKDFETYNNAKILDLNNTATFRLPRETELQTLGNYF 443
Db 380 ---FTSGENNAVGAEY-GTGVFYD-----ETHTKRYGLEVY 413
QY 444 HNEYGKNRFPPEELGLFFD--GPDQDNGL-----YSYLGRKFGDK 480
```

Db 414 YTNADKDTWADYARLSYDRQIGLDNHFQOHCSDGSKYCRPSADKPFY---YKSDR 470
QY 481 GLLPQKSTIVOPAGSQVFNFTFYDAALKDITRLNYSNTVGY-RFGG-----EYTGYYG 534
Db 471 VIYGESHKLLOAAFKKSFDTAKI-----RHNLNVN-LGYDRFGSNLRHODY--YYQ 518
QY 535 SDDEFKRAFGENSPYKHKCNQSCGIYEPVLKYYGKKRA-----NNHVSYSISAD----- 583
Db 519 SAN---RAYSLKTPP-----QNNKKTSPPNGREKNPYWVSIGRGNVVT 559
QY 584 ---FGD-----YMPFA-----SYSRPH----- 598
Db 560 QICLFGNNYTDCTPRSINKSYAAAVNRVLRGWADVAGLRDYRSTHSDGSGYSTG 619
QY 599 -----RMPNIQEMFYSGIGSDGVHTALKPERANTWQ 629
Db 620 THRTLSNAGIVLKPADWLDITYRTSTGFRLPSPAEMYGWRSGDKIKAVKIDPEKSFENKE 679
QY 630 FG-----FNTYKGLLKODDTLGLKLGVYGRIDN-----YIH--- 662
Db 680 AGIVEKGFDFGNLEASWFNNAYRDLIVR-----GYEAQIKDGKEQVKGNPAYLNAQS 730
QY 663 ---NVYGGK-W---WDLNGNIP-SWVSSTGLAYT-IOHRNFKDKVKHKGFELELYD 709
Db 731 ARTGINILKIDWNGYWD---KLPEGWYST--PAYNRVRVDIKKRADTDIQSHL-FD 784
QY 710 YGREFTNLSYAYOKSTQPTNFSNASKEDQLKOGYGLSRVSALPRDYGRLEVGT 769
Db 785 -----AIQPSRYV-----VGSGYDQPEGWGVNGMLT-----YSKAKEIT 819
QY 770 RWLGNKLTLOGAMRYFKGSRATRAERYIDCTNGCNTSNVRQLGKRSIKOTETILAROPLI 829
Db 820 ELLGSRALLNG-----NSRNTKATARTREPWI 847
QY 830 DFVAAVEPKNLIFFRAEVKNLPDRRYI 857
Db 848 VDVSGYTIVKKHFTLRAGVYNLLNHRIV 875

RESULT 4
US-08-478-435-96
; Sequence 96, Application US/08478435
; Patent No. 5922323
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,435
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-462 MIS:Vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-435-96

Query Match 5.7%; Score 278; DB 2; Length 915;
Best Local Similarity 21.4%; Pred. No. 4.4e-15;
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

QY 5 FRKPICFILMGVLMYHHSYAEADAGAGSEAOIQVLEDVHVAK--RVPKDKKVFETDARA 62
Db 7 FRNLICLSLMTAL---PAYAENV-QAG-QAQEKQLDTIOVKAKKQKTRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDQK--SSGIVSLNIRGDSGFGORVNTMVDGIT 118
Db 62 VKTADTLSKEQVLDIRDLTRYDPCIADVVEQGRGASSG---YSIRG-MDKNRVSLVTDLGA 117
QY 119 Q-TFYSTSTDAG--RAGGSSQFGASVDSNFIAGLDVVVKGSFSSAGINSLAGSANLRTLG 175
Db 118 QIOSYTAQAALGCTRTAGSSCAINTEYENVKAVEISKSGNSVEQSGSAGALGAVFQTKT 177
QY 176 VDDVVGNNNTYGLLLK-GLTGTSN--TKGNAMAA-IGARKWL--ESGASGVGLYGHRSRS 229
Db 178 ADVVIGEGROWIQSTAYSGKNRGLTQSTALAGRIGGAELLIRGRHAGEIRAH---- 233
QY 230 VAQNYRVGGGQHGIFNGFAEYLERRKQRYFVQEGGLKFNNSGKWERDFORPYWKTWKYQ 289
Db 234 -----EAAGRGVQSFNRLAPVDDGSKYAYFIVEECK-----NGGHEKCKANP----- 276
QY 290 KYNDPQELQYIEGHDK-----SHRENLAPOYDITPDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KKDVVGEDKROTVSTRDYTGPNRFLA--DPLSYESRSMFLRFGFRFENKRHY 326
QY 340 DGVFNKYTAQ---FRDLNTRKIGSRKII-----NRNYQFNYSLSNLSYANLNL 383
Db 327 IGGILERTQQTFTDRMTVPALTKAVFDANQKQAGSLRNGKYGAGNH-----KYGGL-- 379
QY 384 TAAYNNGRQYKPGSKFTGMLLKDFETYNNAKILDNLNTATFRLPRETELQTTGLGNYF 443
Db 380 ---FTSGENNAPVGAAY-GTGVEYD-----ETHTKSRYGLEYV 413
QY 444 HNEYGNRPPPEELGLFFD--GPQDNGL-----YSYLGREFKGDK 480
Db 414 YTNADKDTWADYARLSYDRQIGLDNHFQOHCSDGSKYCRPSADKPFY---YKSDR 470
QY 481 GLLPQKSTIVOPAGSQVFNFTFYDAALKDITRLNYSNTVGY-RFGG-----EYTGYYG 534
Db 471 VIYGESHKLLOAAFKKSFDTAKI-----RHNLNVN-LGYDRFGSNLRHODY--YYQ 518
QY 535 SDDEFKRAFGENSPYKHKCNQSCGIYEPVLKYYGKKRA-----NNHVSYSISAD----- 583
Db 519 SAN---RAYSLKTPP-----QNNKKTSPPNGREKNPYWVSIGRGNVVT 559
QY 584 ---FGD-----YMPFA-----SYSRPH----- 598

Db 560 QICLFGNNTYDCTPRSRNGSKSYAAVRDNNVRLGRWADVAGAGURDYRSTHSDGVSSTG 619
QY 599 -----RMPNIQEMFYSGIGDSGVHTALKPERANTWQ 629
Db 620 THRTLSNAGIVLKPADWDLTYRTSTGFRLPFAEMYGWRSGDKIKAVKIDPEKSNKE 679
QY 630 FG-----FNTYKGLLKQDDTLGLKLVGYRSIDN-----YIH--- 662
Db 680 AGIVFKGDFGNLEASFNNAYRDLIVR-----GYEAQIKDGKEQVKNPAYLNAQS 730
QY 663 -----NVYCK--W---WDLNGNIP--SWVSSTGLAYT--IOHRNFKDKVHKHGFLELNYD 709
Db 731 ARITGINILKIDWGVND---KLPEGWYST--FAYNRVRVDIKKRDRTDIQSHL-FD 784
QY 710 YGRFETNLVAYOKSTOPTNFSASEPNNAKEDQLKQGYGLSRVSALPRDYGRLEVGT 769
Db 785 -----AIQPSRYV-----VSGGYDQEGKGVNGMLT-----YSKAKEIT 819
QY 770 RWLGNKLTGGAMRYFGKSRATAERYIDGTNGGNTSNVRLQKRSIKOTETLAROPLI 829
Db 820 ELLGSALLNG-----NSRNTKATARTRPWYI 847
QY 830 FDFVAAYPEKKNLIFRAEVKNLFDPRYI 857
Db 848 VDVSYYTVKKHFTLRACVYNLLNHRIV 875

RESULT 5
US-08-337-483-96
; Sequence 96, Application US/08337483
; Patent No. 592562
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 Unviersity Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:1b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-337-483-96

Query Match 5.7%; Score 278; DB 2; Length 915;
Best Local Similarity 21.4%; Pred. No. 4.4e-15;
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;
QY 5 FRLPICFYLGMVLMYHHSYAEADAGRAGSEAOIOVEDVHVAK--RVPDKKKVFTDARA 62
Db 7 FRNLILSLMTAL---PAYAENV-QAG-QAEKQOLDTIOVKAQKQKTRRDNEVYGLGL 61
QY 63 VSTRDIFKSE--NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGVRNMTVDGIT 118
Db 62 VKTADTLSKQOVLDIRDLTRYDPCIAVVEQGRGASSG---YSIRG-WDKNRVSLTVDGLA 117
QY 119 Q-TFYSTSDAG--RAGGSQFASVDSNFTAGLDVVKVGSFSGSAGINSLASLANLTGL 175
Db 118 QIQSYTAAALGGTRTAGSSGAINIEYENVKAVEISKGSNSVQSGSAGALAGSVAFOTKT 177
QY 176 VDVVQGNNTYGLLLK-CLTGTNS--TKGNMAA--IGARKWL--ESGASVCLVGHSPRS 229
Db 178 ADDVIGEGROWGIQSKTAYSGKNRGLTQSIAGALAGRIAGAEALLIRTKRHAGEIRAH--- 233
QY 230 VAQNYRVGGGQHIGNFGAEYLERRKQRYFVQEGGLFNNSGKWERDFQRPYKTKWYQ 289
Db 234 -----EAAGRGVQSFNRLAPVDDGSKYAYFIVEECK-----NGGHECKANP----- 276
QY 290 KYNDPQELQYIEGHDK-----SWRENAPQYDITPIDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KKDVVGEDKQTVSTRDYTPGNRFLA--DPLSYESRSWLFPRCFREENKRYH 326
QY 340 DGVENKYTAQ---PRDLNTKIGSRKII-----NRYQFNYGLSLNSYANLNL 383
Db 327 IGGILERTQOTFTRDMTVPAFLTKAVFDANQKAGSLRGNGKYAGNH-----KYGGL-- 379
QY 384 TAAYNSGRQYPKGSKFTGMLLKDFETYNNNAKTILDNLNTATFRLPRETELQTTLGFNYF 443
Db 380 ---FTSGENNAPVGAAY-GTGVFYD-----RHLNLSV-LGYDRFGSNLRHODY--YYQ 518
QY 444 HNEYGKRNFPPEELGLFED--GPDODNGL-----YSYLRFRFKGDK 480
Db 414 YTNADKDTWADYARLSYDRQIGLDNHFQOOTHCSADGSDKYCRPSADKPFYS---YKSDR 470
QY 481 GLLPQKSTIVQAGSOYFNTFYDAALKDIIYRLNYSTNTVGY-RFGG-----EYTGYY 534
Db 471 VIYGESHKLLQAAFKKSFDTAKI-----RHLSVN-LGYDRFGSNLRHODY--YYQ 518
QY 535 SDDEFKRAFGENSPYKKHCNQSCGIYEPVLKYGKKRA-----NNHSVSISAD----- 583
Db 519 SAN---RAYSLKTPP-----QNNGKKTSPNGREKNPYWVSIGRGNVVTR 559
QY 584 ---FGD-----YFMPEA-----SYSRTH----- 598
Db 560 QICLFGNNTYDCTPRSRNGSKSYAAVRDNNVRLGRWADVAGAGURDYRSTHSDGVSSTG 619
QY 599 -----RMPNIQEMFYSGIGDSGVHTALKPERANTWQ 629
Db 620 THRTLSNAGIVLKPADWDLTYRTSTGFRLPFAEMYGWRSGDKIKAVKIDPEKSNKE 679
QY 630 FG-----FNTYKGLLKQDDTLGLKLVGYRSIDN-----YIH--- 662
Db 680 AGIVFKGDFGNLEASFNNAYRDLIVR-----GYEAQIKDGKEQVKNPAYLNAQS 730
QY 663 -----NVYCK--W---WDLNGNIP--SWVSSTGLAYT--IOHRNFKDKVHKHGFLELNYD 709
Db 731 ARITGINILKIDWGVND---KLPEGWYST--FAYNRVRVDIKKRDRTDIQSHL-FD 784
QY 710 YGRFETNLVAYOKSTOPTNFSASEPNNAKEDQLKQGYGLSRVSALPRDYGRLEVGT 769
Db 785 -----AIQPSRYV-----VSGGYDQEGKGVNGMLT-----YSKAKEIT 819
QY 770 RWLGNKLTGGAMRYFGKSRATAERYIDGTNGGNTSNVRLQKRSIKOTETLAROPLI 829
Db 820 ELLGSALLNG-----NSRNTKATARTRPWYI 847

APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/474,671
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
us-08-474-671-96

Query Match 5.7%; Score 278; DB 3; Length 915;
Best Local Similarity 21.4%; Pred. No. 4.4e-15;
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

QY 5 FRLPICFYLGMVLYHHSVAEDAGRAGSAQIOVLSDVHVAK--RVPKDKKVFETDARA 62
DB 7 FRNLICLSMTAL---PAVAENP-QAG-QAQEKQDITQVAKKQKTRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDDK--SSGIVSINIRGDSGFGRVNTMVDGIT 118
DB 62 VKTADTLSKEQVLDRLTRYDPCGIAVVEQGRGASSG--YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TPTYSTDAG--RAGSSQFASVDSNFTAGLDVYKGFSSGAGINSLAGSANLRTLG 175
DB 118 QIOSYTAQAALGGTRTAGSGAINEIEYENKVAEISKSGNSVBOGSGALAGSVAFTKT 177
QY 176 VDDVVOGNNYTGLLK-GLTGTSN--TKGNAMAA-ICARWL--ESGASVGVLYGHSRS 229
DB 178 ADDVIGEGROWGTSQAYSGKNRGLTQSLALAGRIGCAEALLIRTRHAGEIRAH---- 233
QY 230 VAQNYRVGGGQHGIFGAELERRKORYFVOEGGLAFNSNGKWERDFORPYWKTWKYQ 289
DB 234 -----EAGRGVQSFNRLAPVDDGSGYAYFIVEECK-----NGGHEKCKANP----- 276

QY 290 KYNDPQELQYIEGHDK-----SWRENLAPOYDIPTDPSSLKQOS-----AGNLF--KLEY 339
DB 277 -----KKDVVGEDKRQTVSTRDYTGPNRFLA--DPLSYESRSWLFPRCPGFRENRRHY 326
QY 340 DGVFNKYTAQ---FRDLNLTIGSRKII-----NRYQFNYGLSLNYSANLNL 383
DB 327 IGGILERTQOTFTRDMTVPAFLTKAVFADANQKQAGSLRGNKYGAGNH-----YKGL-- 379
QY 384 TAAVNSGRQYKPGSKFTGMLKDFETYNNAKILDLNNTATFRLPRETELQTTLGFNYF 443
DB 380 ---FTSGENNAPVCAEY-GTGVFYD-----ETHKSRYGLEYV 413
QY 444 HNEYGKNRFPPEELGLFFD--GPDQDNL-----YSYLRGPFKGDK 480
DB 414 YTNADKDTWADYARLSYDRQDGLDNHFQOOTHCSADGSDKYCRPSADKPESY---YKSDR 470
QY 481 GLLPOKSTIVQAGSOYFNTFYDAALKKDIYRLNYSNTVGY-RFGG-----EYTGYYG 534
DB 471 VIYGESHKLQAAFKKSFTAKI-----RHNLNVN-LGYDRFGSNLRHODY--YYQ 518
QY 535 SDEFFKRAFGENSPYKHCNQSCGIYEPVLKYGKKRA-----NNHSVSIASD----- 583
DB 519 SAN---RAYSLKTPP-----QNNGKKTSPNGREKNPYVWSIGRGNVTR 559
QY 584 ---FGD-----YFMPEA-----SYSRTH----- 598
DB 560 QICLFGNNTYDCTPRSINGKSYAAVRDNRVLGRWADVGAGLRDYRSTHSDDGVSSTG 619
QY 599 -----RMPNIOEMVFSOIGDSGVHTALKPERRANTWQ 629
DB 620 THRTLSWNAIVLKPADWLDLTYSSTGFRLPSPAEMYGWRSQDKIKAVKIDPEKSNKE 679
QY 630 FG-----FNTYKGLLKQDDTLGLKLGVYRSRDN-----YIH--- 662
DB 680 AGIVFKGDFGNLEASFNNAYRDLIVR-----GYEAQIKDGKEQVKNPAYLNAQS 730
QY 663 -----NVYCK--W---WDLNGNIP-SWVSSTGLAYT-IQHRNEKDKVHKHGFLELNYD 709
DB 731 ARITGINILKIDWNGVMD---KLPEGWYST--FAYNRVRVDIKKRADRTDIOSHL-FD 784
QY 710 YGREFTNLVAYOKSTQPTNFSADASESPNASKEDQLKQGYGLSRVSALPRDYGRLEVGT 769
DB 785 -----AIQPSRV-----VCSGYDQPEGKGVNGMLT-----YSKAREIT 819
QY 770 RWLGNKLTGAMRYFGKISIRATAEERYIDGTNGNTSNVRQLGKRSTIKQETLAROPLI 829
DB 820 ELLGSALLNG-----NSRNTKATARRTRPWYI 847
QY 830 FDFVAAVEPKKNLIFRAEVKNLFDRIYI 857
DB 848 VDVSGYTVTKKHFTLRAGVYNLLNHRVY 875

RESULT 8
US-08-483-577A-96
; Sequence 96, Application US/08483577A
; Patent No. 6015688
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto

STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,577A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-511
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-577A-96

Query Match 5.7%; Score 278; DB 3; Length 915;
Best local similarity 21.4%; Pred. No. 4.4e-15;
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

QY 5 FRKPCIFYLMGVMYHSHSVAEDRAGSEAOIQVLEDVHVAK--RVPKDKKVFETDARA 62
DB 7 FRMLNLCLSMTAL--PAAVENV-QAG-QAQEKQLDTIQVAKKQKTRDRDNEVTGLGKL 61
QY 63 VSTRQDIFKSSSE-NLDNIVRSIPG-AFTQDQK--SSGIVSLINIRGDSGFRVNTMVDGIT 118
DB 62 VKTADTLKQVLDIRDITRDYDPAIVVEQGRGASSG--YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TPYSTSTAG--RAGSSQFGASVDSNFIAGLDVVYKGSFSGSAGINSLAGSANLRTLG 175
DB 118 QIQSYTAQAALGGTRTAGSGAINIEIYENVKAVEISKSGNSVBOGSGALAGSVAFOTKT 177
QY 176 VDDVVQGNNTYGLLLK-GLTGTSN--TKGNAMAA-IGARKWL--ESGASGVLYLGHSPRS 229
DB 178 ADDVIGSGRWGIQSKTAYSGKNRGLTQSLAGRIGAGAEALLIRTGHRHAGEIRAH---- 233
QY 230 VAQNYRVGGGQHQHIGFAGYLERRKQRYFVQEGGLFNSNSGKWERDFORPYWKTKWYQ 289
DB 234 -----EAAGRCVQSFNRLAPVDDGSKAYFIVEECK-----NGHECKANP----- 276
QY 290 KYNDPQELQKYEIHKD-----SWRENAPQYDITPIDPSLKQOS-----AGNLF--KLEY 339
DB 277 -----KKDVVGEDRQTVSTRDYTGPNRFLA--DPLSYESRSWLFPRGFRFENKRHY 326
QY 340 DGVENKYTAQ---PRDLNLTIGSKHII-----NRNVPNYGLSLNSYANLNL 383
DB 327 IGGILETQOTFTDRDNTVPAFLTKAVFDANQKQAGSLRNGKYAGNH-----KYGL-- 379
QY 384 TAAVNSGRQYKPGSKFTGMLGLKDFETYNNAKILDLNNTATFRLPRETELQTTLGFNYF 443
DB 380 ---FTSGENNAPVGAEE-GTGVFYD-----ETHTKSRYGLEYV 413

QY 444 HNEYKKNRFPPEELGLFFD--GPDODNGL-----XSYLGRFRKGDK 480
DB 414 YTNADKDTWADYARLSYDRQIGLDNHNFOQTHCSADGSKYCRPSADKPSY---YKSDR 470
QY 481 GLLPQKSTIVOPAGSOYFNTFYFDAALKKDIYRLNYSTNTVGY-RFGG-----EYTGYYG 534
DB 471 VIYGESHKLLQAAPKKSFDATAI-----RHNLSVN-LGYDRFGSNLRHODY--YYQ 518
QY 535 SDDEFKRAFGENSPYKKHCNQSGIYEPVLKKYKGRRA-----NNHSVISISAD----- 583
DB 519 SAN---RAYSCLKTTP-----QNNKKKTSNGREKNKPYWISIGRGNVVR 559
QY 584 ---FGD-----YFMPEA-----SYSRTH----- 598
DB 560 QICLFGNNTYDCTPRSINGKSYAAVRDNVRLGRWADVGAGLYDYRSTHSDDGSYSTG 619
QY 599 -----RMPNQEMYFSGOISGDSGVHTALKPRANTWQ 629
DB 620 THRTLWNAGIVLKPADWLDLTYRTSTGFRLPSPAEMYGWRSGDKIKAVKIDPEKSFKE 679
QY 630 FG-----PNTYKKGILLKODDTLGLKLVCYRSIDN-----YIH--- 662
DB 680 AGIVFGDGFNLEASWFFNNAYRDLIVR-----GYEAQIKDGKEQVKGNPAYLNAQS 730
QY 663 -----NVYK--W---WDLNGNIP-SWVSSTGLAYT-IQHRNFKDKVHKHGFLELNYD 709
DB 731 ARITGINILGKIDWNGVWD--KLPEGWYST--FAYNRVRVRIKKRADRTDIQSHL-FD 784
QY 710 YGRFTNLAYOKSTOPTNFSDAESPNNASKEDQLKGGLSRVSALPRDVGRLVGT 769
DB 785 -----AIQPSRYV-----VGSGYDQPEGKVGNGMLT-----YSKAKEIT 819
QY 770 RWLGNKLTGGAMRYFGKSIKIRATAEERYIDGTNGGNTSNVRQLGKRSIKKIKETILAROLI 829
DB 820 ELLGSRALLNG-----NSRNTKATARTTRPYWI 847
QY 830 FDFYAAEPPKKNLIFRAEVKNLFDRIYI 857
DB 848 VDVSGYTVTKKHFTLRAGVYNLLNHRVY 875

RESULT 9
US-08-897-438-96
; Sequence 96, Application US/08897438
; Patent No. 6262016
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,438
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

; SEQUENCE CHARACTERISTICS:	
; LENGTH: 915 amino acids	
; TYPE: amino acid	
; STRANDEDNESS: single	
; TOPOLOGY: linear	
US-08-637-654-96	
Query Match 5.7%; Score 278; DB 4; Length 915;	
Best Local Similarity 21.4%; Pred. No. 4.4e-15;	
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;	
QY	5 FRLKPICFYLMGVMLYHSHYEDRAGRAGSEAOIQVLEDVHVHAK--RVPKDKKVFETDARA 62
Db	7 FRLNLCISLMTAL---PAYAENY-QAG-QAQRKQLODTIQVKAQKQKTRDNEVTGLCKL 61
QY	63 VSTRQDIFKSSSE-NLDNIVRSIPG-AFTQODK---SSGIVSLNIRGDSGFGRVNTWVGIT 118
Db	62 VKTADTSLKEQVLDIRLTRYDPIGVAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117
QY	119 Q-TFYSTSTDAQ--RAGSSQFASVDSNFTAGLDVVKVFGSGSAGINSLAGSANLRTLGL 175
Db	118 QIQSYTAQAALGGTRTAGSGAINIEIYENVKAVEISKGSNSVEQSGSALAGSVAFOQTK 177
QY	176 VDDVVQGNNTYGLLLK-GLTGTNS--TKGNAMAA-IGARKWL--ESGASGVLYGHRRS 229
Db	178 ADDVIGEGRQMGIOSTAYSGKNRGLTQSIAGLAGRIGAEALLIRTGRHAGEIRAH---- 233
QY	230 VAQNYRVGGGQHIGNFGAEYLERRKORYFVQEGGLFNSNGKWERDFQRPYWKTKWYQ 289
Db	234 -----EAAGRGVQSFNRLAPVDDGSKYAYFIVEECK-----NGHEKCANP----- 276
QY	290 KYNDPQELQKYLIEGDK-----SWRENLAPOVDITPIDPSSLKQQS-----AGNLF--KLEY 339
Db	277 -----KKDVGEDKQRTVSTRDYTGPNRFLA--DPLSYESRSWLFPRGFRFENKRHY 326
QY	340 DGVFNKYTAQ---PRDLNLTIGSKII-------NRNYQFNGLSLNSYANLNL 383
Db	327 IGGILERTQOTFDRTMTVPFLTKAVFDANQKQAGSLRGNGKYAGNH-----KYGL-- 379
QY	384 TAAVNSGRQYKPKSGFTGMLLKDFETYNNAKILDLNLTATFRLPRETELQTTLGFNYF 443
Db	380 ---FTSGENNAVPVGAEV-GTGVFYD-----ETHKSRYGLEYV 413
QY	444 HNEYGKNRFPPELGLFED--GPDQDNGL-----YSYLGRFRKGDK 480
Db	414 YTNADKDTWADYARLSYDROGIGLDNHFQOTHCSADGSKYCRPSADKPFYSY---YKSDR 470
QY	481 GLLPKQSTIVQAGSQVFNFTFYDAALKKDIYRLNYSNTVGY--RFGG-----EYTCYIG 534
Db	471 VIYGESHKLLQAPKKSFDTAI-----RHLNLSVN-LGVDRFGSNLHQDY--YYQ 518
QY	535 SDDEFFKRAFGENSPYKKHCKNQSGGIYEPVLKKYKKRA-----NNHSVYSISAD----- 583
Db	519 SAN---RAYSLKTPP-----QNNKKTSNGREKNFYWYISGRGNVVR 559
QY	584 ---FGD-----YFMPPA-----SYSRTH----- 598
Db	560 QICLFGNNTYDCTPRSINGKSYAAVRNVLGRWADVAGLRYDYRSTHSDGDSVSTG 619
QY	599 -----RMPNIQEMVFSQIGDSGVHTALKPERANTWQ 629
Db	620 THRTLSWAGIVLKPADWLDLTYRTSGFRLPSPFAEMYGWRSGDKIKAVKIDPEKSENKE 679
QY	630 FG-----FNTYKKGILLKODDTLGLKLVGYSRSDN-----YIH--- 662
Db	680 AGIVFKGDFCNLEASWNNAYRLIVR-----GYEAQIKDGKEQVKGNPAYLNAQS 730
QY	663 -----NVYQK--W---WDLGNIP--SWSSSTGLAYT-IQHRNFKDKVHKHGFELNLYD 709
Db	731 ARITGINILKIDWNGVWD---KLPEGWYST--FAYNRVRVRDIKKRADRTDIQSHL-FD 784
QY	710 YGREFTNLAYQKSTQPTNFSDASESPNNAKEDQLKQGYGLSRVSALPRDYGRLEVGT 769
Db	785 -----AIQPSRYV-----VSGYDQPGKGVNGMLT-----YSKAKEIT 819
QY	770 RWLGNKLTLCGAMRYFGKSIIRATAEERVIDGTNGNTSNVRQLGRSRIKQITETLARQPLI 829
Db	820 ELLGSRALLNG-----NSRNTKATARRTRPWYI 847
QY	830 FDFVAAAYEPKKNLIFRAEVKNLFDRIYI 857
Db	848 VDVSGYITVKKHFTLRACGVYNLLNHRIV 875
RESULT 11	
US-08-487-890A-94	
; Sequence 94, Application US/08487890A	
; Patent No. 5708149	
; GENERAL INFORMATION:	
; APPLICANT: Loosmore, Sheena	
; APPLICANT: Harkness, Robin	
; APPLICANT: Schryvers, Anthony	
; APPLICANT: Chong, Pele	
; APPLICANT: Gray-Owen, Scott	
; APPLICANT: Yang, Yan-ping	
; APPLICANT: Murdin, Andrew	
; APPLICANT: Klein, Michel	
; TITLE OF INVENTION: Transferrin Receptor Genes	
; NUMBER OF SEQUENCES: 147	
; CORRESPONDENCE ADDRESS:	
; ADDRESSEE: Sim & McBurney	
; STREET: 6th Floor, 330 University Avenue	
; CITY: Toronto	
; STATE: Ontario	
; COUNTRY: Canada	
; ZIP: M5G 1R7	
; COMPUTER READABLE FORM:	
; COMPUTER: IBM PC compatible	
; OPERATING SYSTEM: PC-DOS/MS-DOS	
; SOFTWARE: Patent In Release #1.0, Version #1.25	
; CURRENT APPLICATION DATA:	
; APPLICATION NUMBER: US/08/487,890A	
; FILING DATE: 07-JUN-1993	
; CLASSIFICATION: 435	
; PRIOR APPLICATION DATA:	
; APPLICATION NUMBER: US 08/175,116	
; FILING DATE: 29-DEC-1993	
; PRIOR APPLICATION DATA:	
; APPLICATION NUMBER: US 08/148,968	
; FILING DATE: 08-NOV-1993	
; ATTORNEY/AGENT INFORMATION:	
; NAME: Stewart, Michael I	
; REGISTRATION NUMBER: 24,973	
; REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb	
; TELECOMMUNICATION INFORMATION:	
; TELEPHONE: (416) 595-1155	
; TELEFAX: (416) 595-1163	
; INFORMATION FOR SEQ ID NO: 94:	
; SEQUENCE CHARACTERISTICS:	
; LENGTH: 908 amino acids	
; TYPE: amino acid	
; STRANDEDNESS: single	
; TOPOLOGY: linear	
US-08-487-890A-94	
Query Match 5.5%; Score 269.5; DB 1; Length 908;	
Best Local Similarity 21.0%; Pred. No. 2.3e-14;	
Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;	
QY	5 FRLKPICFYLMGVMLYHSHYEDAGRAGSEAOIQVLEDVHVHAK--RVPKDKKVFETDARA 62
Db	7 FRLNLCISLMTAL---PYVAENV--QAEQAQEKQDITQVKAQKQKTRDNEVTGLCKL 61

QY 63 VSTRQDIFKSSE-NLDNIVRSIPG-AFTQDK--SSGIVSLNIRGDSGFRVNTWVDGIT 118
Db 62 VKSDTLKSKQVLRDLTRDIPGIAVVEQGRASS--YSIRG-MDKNRVSLTVDGV 117
QY 119 Q-TFYSTSDAG--RAGGSQFQSGVDSNFIAGLDVVKFSGFSAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEYENKVAIEISKSGNSSEYNGALAGSVAFOK 177
QY 176 VDDVVOGNNYTGLLK-GLTGTNTSGKNAMAAGARKWLESAGSVGLYGHSSRSVAQNY 234
Db 178 AADIIGBKQWGIQSKTAYSGKDHQSLALAGR---SGGAELLYTKRRGREIHAH 233
QY 235 RVGGGQHIGNFAEXLERKQ-----RYFVQE---GGLKFSN-----SCKWERDF 278
Db 234 KDAKGG--VQSFNRLVDEKKGSGRYRIFVEEHCNGYACKNKLKEDASVKDER-- 289
QY 279 QRPYKTKWKYQKNDPOE-LQKIEGHDKSWRENLAPOYDITPIDPSLQKQAGNLFKL 337
Db 290 ----KTVSTQDYTSNRLANLEYSQSWL--FRGW-----HLDNR 326
QY 338 EYDGVFNKYTAQFDLNTKIGSKRIINRYOFNGLSLNYSANLNTAAVNSGRQKYPK 396
Db 327 HYGVAVLERTQOTPDTRDMTPVAYFTSDEYVPGSLKGLGY-----SGDNKAERL 376
QY 397 ----GSKFTGWGLLKDFETYNNAKILDLNLTATFRLPRETELQTTGLFPNYFNEYCKNR 451
Db 377 FVOGEGTLOGIG-----YGTGVFYD-----ERHTKNRYGVYVYHNADKDT 418
QY 452 FPELGLFFD--GPDQDNGL-----YSYLGFRKGDGKLLPKST 488
Db 419 WADYARLSYDROGLDLNRLQOHCSDGDKNCRPDGKNPYSE---YKSDRIYEESRN 475
QY 489 IVOPAGSQYNT-----FYDAAAL----- 507
Db 476 LFOAVFKKAFDTAKIRHNLSTNLGYDRFKSQLSHSDYILQNAVOAYDLITPKKPPFNGS 535
QY 508 KDIYRLNYSNTVG---YRGE-YT-----GYGSDDEKFRAGFENSPYKK 552
Db 536 KDNPIRVISGKTNTVSPICRFGNTTDTCPRNIGNGYI-----AAVDNVRCLR 587
QY 553 HCNQSGI-YEPVLKYKKRANNSHS-----ISADFGDYFMPFASYRTH-----R 599
Db 588 WADVGAGIRYD----YRSTHSEDKSVSTGTHRNLWNAGVVLKPTWMDLTYRSTGFR 642
QY 600 MPNTQEMWFSQIGSDGVHTALKPERANTWQFG-----PNTYKGLLKKODDT 645
Db 643 LPSPAEMYWRAGBSLTKLDLKPEKSFNREAGIVFKGDFGNLEASYFNAYRDLI----- 697
QY 646 LGLKLVYRSRIDN-----YIH-----NRYGK--WWDLNGNIPSHVSSITGL 684
Db 698 ----AFGYETRTQNGQTSASGDPGYRNAQNAIAGINILGIDKHGVMGGLPDGLYST-L 752
QY 685 AY-----TIQHRNFKDKVHKHFELELVYD--GFFETNLSVAYOKSTQPT 728
Db 753 AYNRKIKVKDADRTRFVTSYLFDAVQPSRYVLGLGYDHPDGIWGTINTWFTYSK----- 807
QY 729 NFSDASEPNNAKEDOLKQCYGLSRVSALPRDYGRLEVTRWLGKNTLTGGAMRYECKS 788
Db 808 ----AKSVDEL-----LGSQALLNG----- 823
QY 789 IRATAERYIDGTNGGNTSVNRQLGKRSIKOTETLARQPLIFDYAAYEPKKNLIFRAEV 848
Db 824 -----NANAKAASRR-----TRPWYVTVDSGYNTYNTKKH/LTLRAGV 859
QY 849 KNLFDRIYI 857
Db 860 YNLLNYRYV 868

RESULT 12

us-08-478-435-94

; Sequence 94, Application US/08478435

; Patent No. 5922323

GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,435
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-478-435-94

Query Match 5.5%; Score 269.5; DB 2; Length 908;

Best Local Similarity 21.0%; Pred. No. 2:3e-14;
Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;QY 5 FRLPFCIFLGMVLMYHHSYAEDAGRAGSEAIQVLEDVHVYKAK--RVPKDKKVFETDARA 62
Db 7 FRLNLCLEMTAL--PVYAENV--QAEQAEKQLDTIQVAKKQKTRDRNEVTGLGKL 61

QY 63 VSTRQDIFKSSE-NLDNIVRSIPG-AFTQDK--SSGIVSLNIRGDSGFRVNTWVDGIT 118

Db 62 VKSDTLKSKQVLRDLTRDIPGIAVVEQGRASSG---YSIRG-MDKNRVSLTVDGV 117

QY 119 Q-TFYSTSDAG--RAGGSQFQSGVDSNFIAGLDVVKFSGFSAGINSLAGSANLRTLG 175

Db 118 QIQSYTAQAALGGTRTAGSSGAINIEYENKVAIEISKSGNSSEYNGALAGSVAFOK 177

QY 176 VDDVVOGNNYTGLLK-GLTGTNTSGKNAMAAGARKWLESAGSVGLYGHSSRSVAQNY 234

Db 178 AADIIGBKQWGIQSKTAYSGKDHQSLALAGR---SGGAELLYTKRRGREIHAH 233

Db 643 LPSFAEMYGWRAGSLKTLDLKPEKSFNREAGIVFKGDFGNLEASYNFNNAYRDLI----- 697
QY 646 LGLKLVGRSRIDN-----YIH-----NVYK--WMDLNGNIPSWVSSTGL 684
Db 698 ----AFGYETQNGQTSASCDPCYRNAQNAIAGINILKIDHWGVWGLPDGLYST-L 752
QY 685 AY-----TIOHRNFKDKVHKHGFELNLDY--GRFFTNLSYAYOKSTQPT 728
Db 753 AYNRIKVKDADIRADRTFVTSYLFDAVQPSRYVLGLOYDHPDGIWINTWFTYSK----- 807
QY 729 NFSDASESPNNASKEDQLKQGYGLSRYSALPRDYGRLEVGRWLGKNTLGGAMRYEFGS 788
Db 808 -----AKSVDEL-----LGSQALLNG----- 823
QY 789 IRATAEERYIDGTNGGNTSVNROLGKRSIKQETETLARQPLIFDFYAAEYEPKKNLIFRAEV 848
Db 824 -----NANAKKAASRR-----TRPWVTDVSGYINIKKHLTLRAGV 859
QY 849 KNLFDRRYI 857
Db 860 YNLLNRYV 868

RESULT 15

us-08-474-671-94
; Sequence 94, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-474-671-94

Query Match 5.5%; Score 269.5; DB 3; Length 908;
Best Local Similarity 21.0%; Pred. No. 2.3e-14;
Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;

QY 5 FRLKPTCYFLMGVNLVHSHYAEDAGRAGSEAQIQVLEVDVHKAK--RVPKDKKVFDTDARA 62
Db 7 FRNLILCLSLMTAL---PYVAENV--OAEQAOEKOLDIQVKAKKOKTRRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE--NLDNIVRSIPG--AFTQODK--SSGIVSLNIRGDSGFCRVNTWVDGIT 118
Db 62 VKSSDLSKEQVLNIRDLTRYDFGIAVVEQGRGASS---YSIRG--MDKNRVSLTYDGV 117
QY 119 Q-TFYSTSTDAG--RAGGSSQFGASVDSNFIAGLDVVVKGSFGSAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEIYENVKAVEISKGSNSSEYNGALAGSVAFTKT 177
QY 176 VDDVQGNNTYGLLLK--GLTGTNSTKGNMAAIGARKWLESASVGVLYGHSSRSVAQNY 234
Db 178 AADIIGEGKQWGIQSKTAYSGKDHALTQSLALAGR---SGGAELLIYTKRGRRETHAH 233
QY 235 RVGGGGQHIGNFGAEYLERKQ-----RYFVQ--GGLKFSN-----SGKWERDF 278
Db 234 KDAGK--VQSFNRLVDEDKKEGGSQRYFIVEEENHNGYAAACKNKLKEDASVKDER-- 289
QY 279 QRPYKTKYQKYNDOE--LQKYIEGHDKSWRENLAPOYDITPIDPSSLKQQSAGNLFKL 337
Db 290 -----KTVSTQDYTGSNRLLANPLEYGSQSWL--FRGW-----HLDNR 326
QY 338 EYDGVFNKYTAQFRDLNLTIGSKRIINRNYQFNYGSLNSYANLNLTAAYNSGRQKYPK- 396
Db 327 HYVGAVLERTQQTFTDRMTVPAYFTSEYVPGSLKGLGY-----SGDNKAERL 376
QY 397 -----GSKFTGWGLLKDFETYNNAKILDLNLTATFRLPRETELQTLTGFWFNEYKCN 451
Db 377 FVQEGSTLOGIG-----YGTGVFYD-----ERHTKNRYGVYVYHNADKDT 418
QY 452 FPEELGLFFD--GPDODNGL-----YSYLGFRFGDKGLLPKST 488
Db 419 WADYARLSYDRQIGDLDLNLQOQTHCSHDGSKNCRPDGNKPYSF---YKSDRMIEBSRN 475
QY 489 IVOPAGSQYFNT-----FYDAAAL----- 507
Db 476 LFOAVFKKAFDTAKIRHNLNLSINLGYDRFKSQLSHSDYLLQNAVQAYDLITPKPPFPNGS 535
QY 508 KKDIYRLNYSNTVG---YRFGE-YT-----GYGSDDEPKRAFGENSPYKK 552
Db 536 KNPYRVISGKTTVNTSPICRFGNNTYDCTPRNIGNGY-----AAVDNVLGR 587
QY 553 HCNQSCGI-YEPVLKYYKKRANNHSVS-----ISADFGDYMFPFASYSRTH-----R 599
Db 588 WADVGAGIRYD---YRSTHSEDKSVSTCTHRNLSNAGVWLKPTWMDLTYRSTGFR 642
QY 600 MPNIQMYFSQIGDSGVHTALKPERANTWQF-----FNTYKGLLLKDDT 645
Db 643 LPSFAEMYGWRAGESLKTLDLKPESFNREAGIVFKGDFGNLEASYNFNNAYRDLI----- 697
QY 646 LGLKLVGRSRIDN-----YIH-----NVYK--WMDLNGNIPSWVSSTGL 684
Db 698 ----AFGYETRTONGQTSASGDCPYRNAQNAIAGINILKIDHWGVWGLPDGLYST-L 752
QY 685 AY-----TIOHRNFKDKVHKHGFELNLDY--GRFFTNLSYAYOKSTQPT 728
Db 753 AYNRIKVKDADIRADRTFVTSYLFDAVQPSRYVLGLOYDHPDGIWINTWFTYSK----- 807
QY 729 NFSDASESPNNASKEDQLKQGYGLSRYSALPRDYGRLEVGRWLGKNTLGGAMRYEFGS 788

Db 808 -----AKSVDEL-----LGSOLLNG----- 823
QY 789 IRATAERYIDGTNGGNTSNVQLGKRSIKOTETLARQPLIFDFYAAAYEPKKNLIFRAEV 848
Db 824 -----NANAKAASRR-----TRPWYVTDVSGYINIKKHLTLRAGV 859
QY 849 KNLFDREYI 857
Db 860 YNLLNYRYV 868

Search completed: July 24, 2002, 08:53:48
Job time: 520 sec


```
Qy 61 RAVSTRQDIFKSSNLDNIVRSIPGAFTQDDKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Db 61 RAVSTRQDIFKSSNLDNIVRSIPGAFTQDDKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Qy 121 FYSTSTDAGRAGSSQFASVDSNFIAGLDVVKGVSFSGSAGINSLAGSANLRTLGVDVV 180
Db 121 FYSTSTDAGRAGSSQFASVDSNFIAGLDVVKGVSFSGSAGINSLAGSANLRTLGVDVV 180
Qy 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYIGHSRSSVAQNTYRVGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYIGHSRSSVAQNTYRVGGG 240
Qy 241 OHIGNFAEYLERKQRYFVQEGGLKFNSNGKWERDFQRPYKTKWYQKYNQDPELOKY 300
Db 241 OHIGNFAEYLERKQRYFVQEGGLKFNSNGKWERDFQRPYKTKWYQKYNQDPELOKY 300
Qy 301 IEHDKSWRENLAPOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQPRDLNKTGSR 360
Db 301 IEHDKSWRENLAPOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQPRDLNKTGSR 360
Qy 361 KIINRNTQFNYGLSLNSYANLNTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILD 420
Db 361 KIINRNTQFNYGLSLNSYANLNTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILD 420
Qy 421 NNTATFRLPRETELOTTLGFNFYFNEHYGKNRFPPEELGLFDPGPDQDNGLSYLGFRFGDK 480
Db 421 NNTATFRLPRETELOTTLGFNFYFNEHYGKNRFPPEELGLFDPGPDQDNGLSYLGFRFGDK 480
Qy 481 GLLPQKSTIVOPAGSQVFNTFYFDAALKKDIYRLNYSNTVGYRFGGEYTYGSDDEFK 540
Db 481 GLLPQKSTIVOPAGSQVFNTFYFDAALKKDIYRLNYSNTVGYRFGGEYTYGSDDEFK 540
Qy 541 RAFGENSPTYKKHCNQSCGIYEPVLYKKGKRRANNHSSVSIADFGDYFMPFASYSRTHRM 600
Db 541 RAFGENSPTYKKHCNQSCGIYEPVLYKKGKRRANNHSSVSIADFGDYFMPFASYSRTHRM 600
Qy 601 PNIQEMFYSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
Db 601 PNIQEMFYSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
Qy 661 IHNVYKRWDLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNLYDYGRFPTNLSYA 720
Db 661 IHNVYKRWDLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNLYDYGRFPTNLSYA 720
Qy 721 YOKSTOPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTGG 780
Db 721 YOKSTOPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTGG 780
Qy 781 AMRYFGKSIRATAEERYIDGTNGNTSNVRLGKRSIKQETTLARQPLIFDFYAAEPKK 840
Db 781 AMRYFGKSIRATAEERYIDGTNGNTSNVRLGKRSIKQETTLARQPLIFDFYAAEPKK 840
Qy 901 SKSVLTNFARGRTFLITMSYKF 922
Db 901 SKSVLTNFARGRTFLITMSYKF 922
```

RESULT 2

```
US-09-762-926-2
; Sequence 2, Application us/09762926
; GENERAL INFORMATION:
; APPLICANT: Thonnard, Joelle
; TITLE OF INVENTION: Novel Compounds
; FILE REFERENCE: BM45330
; CURRENT APPLICATION NUMBER: US/09/762, 926
; PRIOR FILING DATE: 2001-02-14
; CURRENT FILING DATE: PCT/EP99/05989
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: GB 9818004.5
```

```
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Bacteria
US-09-762-926-2
```

```
Query Match 99.88; Score 4894; DB 21; Length 922;
Best Local Similarity 99.88; Pred. No. 0;
Matches 920; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 MRSSFRLKPTCFYLMGVMLYHHSVAEDAGRAGSEAOQLVLEHVHAKRVPKDKKVFDTA 60
Db 1 MRSSFRLKPTCFYLMGVMLYHHSVAEDAGRAGSEAOQLVLEHVHAKRVPKDKKVFDTA 60
Qy 61 RAVSTRQDIFKSSNLDNIVRSIPGAFTQDDKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Db 61 RAVSTRQDIFKSSNLDNIVRSIPGAFTQDDKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Qy 121 FYSTSTDAGRAGSSQFASVDSNFIAGLDVVKGVSFSGSAGINSLAGSANLRTLGVDVV 180
Db 121 FYSTSTDAGRAGSSQFASVDSNFIAGLDVVKGVSFSGSAGINSLAGSANLRTLGVDVV 180
Qy 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYIGHSRSSVAQNTYRVGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYIGHSRSSVAQNTYRVGGG 240
Qy 241 OHIGNFAEYLERKQRYFVQEGGLKFNSNGKWERDFQRPYKTKWYQKYNQDPELOKY 300
Db 241 OHIGNFAEYLERKQRYFVQEGGLKFNSNGKWERDFQRPYKTKWYQKYNQDPELOKY 300
Qy 301 IEHDKSWRENLAPOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQPRDLNKTGSR 360
Db 301 IEHDKSWRENLAPOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQPRDLNKTGSR 360
Qy 361 KIINRNTQFNYGLSLNSYANLNTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILD 420
Db 361 KIINRNTQFNYGLSLNSYANLNTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILD 420
Qy 421 NNTATFRLPRETELOTTLGFNFYFNEHYGKNRFPPEELGLFDPGPDQDNGLSYLGFRFGDK 480
Db 421 NNTATFRLPRETELOTTLGFNFYFNEHYGKNRFPPEELGLFDPGPDQDNGLSYLGFRFGDK 480
Qy 481 GLLPQKSTIVOPAGSQVFNTFYFDAALKKDIYRLNYSNTVGYRFGGEYTYGSDDEFK 540
Db 481 GLLPQKSTIVOPAGSQVFNTFYFDAALKKDIYRLNYSNTVGYRFGGEYTYGSDDEFK 540
Qy 541 RAFGENSPTYKKHCNQSCGIYEPVLYKKGKRRANNHSSVSIADFGDYFMPFASYSRTHRM 600
Db 541 RAFGENSPTYKKHCNQSCGIYEPVLYKKGKRRANNHSSVSIADFGDYFMPFASYSRTHRM 600
Qy 601 PNIQEMFYSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
Db 601 PNIQEMFYSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
Qy 661 IHNVYKRWDLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNLYDYGRFPTNLSYA 720
Db 661 IHNVYKRWDLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNLYDYGRFPTNLSYA 720
Qy 721 YOKSTOPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTGG 780
Db 721 YOKSTOPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTGG 780
Qy 781 AMRYFGKSIRATAEERYIDGTNGNTSNVRLGKRSIKQETTLARQPLIFDFYAAEPKK 840
Db 781 AMRYFGKSIRATAEERYIDGTNGNTSNVRLGKRSIKQETTLARQPLIFDFYAAEPKK 840
Qy 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFDPKDDEEVTGNADKTLNCKYGGT 900
Db 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFDPKDDEEVTGNADKTLNCKYGGT 900
```

QY 901 SKSVLTNFARGRTFLITMSYKF 922
Db 901 SKSVLTNFARGRTFLITMSYKF 922
RESULT 3
US-09-762-926-6
; Sequence 6, Application US/09762926
; GENERAL INFORMATION:
; APPLICANT: Thonnard, Joelle
; TITLE OF INVENTION: Novel Compounds
; FILE REFERENCE: BM45330
; CURRENT APPLICATION NUMBER: US/09/762,926
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: PCT/EP99/05989
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: GB 9818004.5
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 921
; TYPE: PRT
; ORGANISM: Bacteria
US-09-762-926-6

Query Match 97.5%; Score 4779.5; DB 21; Length 921;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 903; Conservative 7; Mismatches 11; Indels 1; Gaps 1;
QY 1 MRSFRLKPCFCYLMGYLHSHYAEADAGRAGSEAIQVLEDVHVAKRVPKDKKVFDTA 60
Db 1 MRSFRLKPCFCYLMGYLHSHYAEADAGRAGSEAIQVLEDVHVAKRVPKDKKVFDTA 60
QY 61 RAVSTRODIPKSSNLDNIIVRSIPGAFTQODKSSGIVSLNIRGDSGGRVNTWVDGITOT 120
Db 61 RAVSTRODIPKSSNLDNIIVRSIPGAFTQODKSSGIVSLNIRGDSGGRVNTWVDGITOT 120
QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTILGVDVV 180
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTILGVDVV 180
QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGG 240
QY 241 OHIGNFCAEYLERRKORYFVQEGGLKFNNSGKWERDQRPYKTKWYKYNDOPELOKY 300
Db 241 OHIGNFCAEYLERRKORYFVQEGGLKFNNSGKWERDQRPYKTKWYKYNDOPELOKY 300
QY 301 IEHDKSWRENLAPOYDITPIDPSSLKQSQSAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360
Db 301 IEHDKSWRENLAPOYDITPIDPSSLKQSQSAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360
QY 360 KIINRNTQFYGLSLNLTAAVNSGRQYPKGSKFTGWGLLKDFETYNNAKLTLDL 419
Db 360 KIINRNTQFYGLSLNLTAAVNSGRQYPKGSKFTGWGLLKDFETYNNAKLTLDL 419
QY 421 NNTATFRLPRETLOTTFLGFNHYNEYGKRFPEELGLFFDGDPODNGLYSYLGRPKGDK 480
Db 421 NNTATFRLPRETLOTTFLGFNHYNEYGKRFPEELGLFFDGDPODNGLYSYLGRPKGDK 480
QY 481 GLLPQKSTIVQAGSOYFNTFYDAAALKDIIYRLNYSNTVYRFGGEYTYGYSDDDEFK 540
Db 481 GLLPQKSTIVQAGSOYFNTFYDAAALKDIIYRLNYSNTVYRFGGEYTYGYSDDDEFK 540
QY 541 RAFGENSPYKKHCNCSGCIYEPVLKYKGRANNHNSVTSADFQDFMPFASYSRTHRM 600
Db 541 RAFGENSPYKKHCNCSGCIYEPVLKYKGRANNHNSVTSADFQDFMPFASYSRTHRM 600
QY 601 PNIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLGVYRSRIDNY 660

Db 600 PNIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLGVYRSRIDNY 659
QY 661 IHNVYKQWDLNGNIPSWVSSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFFTNLSYA 720
Db 660 IHNVYKQWDLNGNIPSWVSSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFFTNLSYA 719
QY 721 YKSTQPTNFSASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEFVGTWGLGNKLTIGG 780
Db 720 YKSTQPTNFSASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEFVGTWGLGNKLTIGG 779
QY 781 AMRYFGKSIATAEERYIDCTNGGNTSNVROLGKRSIKQETLARQPLIFDYAAYEPKK 840
Db 780 AMRYFGKSIATAEERYIDCTNGGNTSNVROLGKRSIKQETLARQPLIFDYAAYEPKK 839
QY 841 NLIFRAEVKNLFFORRYIDPLDAGNDAATQRYSSDFDKDEEVTCTNADKTLCKNGYGGT 900
Db 840 NLIFRAEVKNLFFORRYIDPLDAGNDAATQRYSSDFDKDEEVTCTNADKTLCKNGYGGT 899
QY 901 SKSVLTNFARGRTFLITMSYKF 922
Db 900 SKSVLTNFARGRTFLITMSYKF 921

RESULT 4

US-09-303-518D-884
; Sequence 884, Application US/09303518D
; GENERAL INFORMATION:
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisserial Antigens
; FILE REFERENCE: CHIR0160
; CURRENT APPLICATION NUMBER: US/09/303,518D
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 884
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-09-303-518D-884

Query Match 96.9%; Score 4754; DB 17; Length 922;
Best Local Similarity 96.3%; Pred. No. 0;
Matches 888; Conservative 16; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRSFRLKPCFCYLMGYLHSHYAEADAGRAGSEAIQVLEDVHVAKRVPKDKKVFDTA 60
Db 1 MRSFRLKPCFCYLMGYLHSHYAEADAGRAGSEAIQVLEDVHVAKRVPKDKKVFDTA 60
QY 61 RAVSTRODIPKSSNLDNIIVRSIPGAFTQODKSSGIVSLNIRGDSGGRVNTWVDGITOT 120
Db 61 RAVSTRODIPKSSNLDNIIVRSIPGAFTQODKSSGIVSLNIRGDSGGRVNTWVDGITOT 120
QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTILGVDVV 180
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTILGVDVV 180
QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGG 240
QY 241 OHIGNFCAEYLERRKORYFVQEGGLKFNNSGKWERDQRPYKTKWYKYNDOPELOKY 300
Db 241 OHIGNFCAEYLERRKORYFVQEGGLKFNNSGKWERDQRPYKTKWYKYNDOPELOKY 300
QY 301 IEHDKSWRENLAPOYDITPIDPSSLKQSQSAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360
Db 301 IEHDKSWRENLAPOYDITPIDPSSLKQSQSAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360


```

; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 878
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (279)..(279)
; OTHER INFORMATION: Xaa= any amino acid
US-09-303-518D-878

Query Match      93.8%; Score 4599.5; DB 17; Length 888;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 869; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

Qy 34 EAQIOVLEHVHAKRVPKDKVFTDARAVSTRODIFKSSSENLDNIVRSIPGAFTQODKS 93
Db 1 EAQIOVLEHVHAKRVPKDKVFTDARAVSTRODIFKSSSENLDNIVRSIPGAFTQODKS 60

Qy 94 SGIVSLNIRGDSGFRVNTMVDGTTQTFYSTDAGRAGSSQFGASVDSNFIAGLDVVK 153
Db 61 SGIVSLNIRGDSGFRVNTMVDGTTQTFYSTDAGRAGSSQFGASVDSNFIAGLDVVK 120

Qy 154 GSPSGSAGINSLAGSANLRILGVDVVQGNNTYGLLLKGLTGNTSTKGNMAAIGARKWL 213
Db 121 GSPSGSAGINSLAGSANLRILGVDVVQGNNTYGLLLKGLTGNTSTKGNMAAIGARKWL 180

Qy 214 ESGASVGLVGHSSRSVAQNYRVGGGGOHIGNFGAEYLERRKORYFVOEGGLFNSNGK 273
Db 181 ESGASVGLVGHSSRSVAQNYRVGGGGOHIGNFGAEYLERRKORYFVOEGALKFNSDSGK 240

Qy 274 WERDFQRYWTKWKYQKYNQDPOELQKYIEGDKSWRENLAPOYDITPIDPSSLKQOQAGN 333
Db 241 WERDLQWQWKYKPYKNYN-QELQKYIEHDKSWRENLAPOYDITPIDPSSLKQOQAGN 299

Qy 334 LFKLEYDVFNKTYAOPRDNLTKGSRKIIIRNRYQFNYSGLSANYLNLTAAVNSGRQK 393
Db 300 LFKLEYDVFNKTYAOPRDNLTKGSRKIIIRNRYQFNYSGLSANYLNLTAAVNSGRQK 359

Qy 394 YPKGSKFTGMLLKDFEYTNNAKILDLNNTATPLPRETELQTLTGPNFYHNEYGKNRFP 453
Db 360 YPKGSKFTGMLLKDFEYTNNAKILDLNNTATPLPRETELQTLTGPNFYHNEYGKNRFP 419

Qy 454 EELGLFPDGDQDNGLYSILGRFGDKGLLPQKSTIVOPAGSQYFNTFYDAALKKDIYR 513
Db 420 EELGLFPDGDQDNGLYSILGRFGDKGLLPQKSTIVOPAGSQYFNTFYDAALKKDIYR 479

Qy 514 LNYSTNTVGYRFGGEXTGYGSDDEFRAGFGENSPYKKHCNQCGLIYEPVLKKYGGKRA 573
Db 480 LNYSTNTVGYRFGGEXTGYGSDDEFRAGFGENSPYKKHCNQCGLIYEPVLKKYGGKRA 539

Qy 574 NNHSVSIADFGDYFMPFASYSTRHPNPTQEMFYFSGIDSGVHTALKPERANTWQGFN 633
Db 540 NNHSVSIADFGDYFMPFASYSTRHPNPTQEMFYFSGIDSGVHTALKPERANTWQGFN 599

Qy 634 TYKKGLLKQDDTLGLKLVGRYSRIDNTYIHNVYKQWDLNPNISWSSTGLAYTIQHRNF 693
Db 600 TYKKGLLKQDDTLGLKLVGRYSRIDNTYIHNVYKQWDLNPNISWSSTGLAYTIQHRNF 659

Qy 694 KDKVHKHGFLELNDYDGRFFTNLSAYOKSTOPTNFSDASESPNNASKEDQLKQGYLS 753
Db 660 KDKVHKHGFLELNDYDGRFFTNLSAYOKSTOPTNFSDASESPNNASKEDQLKQGYLS 719

Qy 754 RVSAIPRDYGRLEVGTGRLGNKLTGLGAMRYFGKSRATAEERYIDGTNGNTSNVROLG 813
Db 720 RVSAIPRDYGRLEVGTGRLGNKLTGLGAMRYFGKSRATAEERYIDGTNGNTSNFROLG 779

Qy 814 KRSIKQETLARQPLIFDFYAAVEPKKNLIFRAEVKNLFDRIYDPLDAGNDAATQRYYS 873
Db 780 KRSIKQETLARQPLIFDFYAAVEPKKNLIFRAEVKNLFDRIYDPLDAGNDAATQRYYS 839
```

874 SFDPKDKDEVTNADKTLNCKYGGTSKSVLTNFAARGRTFLMTSYKF 922
880 SFDPKDKDEVTNADKTLNCKYGGTSKSVLTNFAARGRTFLMTSYKF 888

7
US-09-303-518D-880
; Sequence 880, Application US/09303518D
; GENERAL INFORMATION:
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisserial Antigens
; FILE REFERENCE: CHIR0160
; CURRENT APPLICATION NUMBER: US/09/303,518D
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 880
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (22)..(22)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (38)..(38)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (40)..(40)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (45)..(45)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (52)..(52)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (68)..(68)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (95)..(95)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (124)..(124)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (133)..(133)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (277)..(277)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (355)..(355)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (370)..(370)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (414)..(414)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (477)..(477)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (483)..(483)
; OTHER INFORMATION: Xaa= any amino acid

[illegible][illegible]

```
Db 407 VLAHNVGQYKPKGSTFTGWLKDKDFETKNTANLFDLNNHTFNLPKQMDLTTTGLNI 466
Qy 443 FHNEYGNRRPEELGDFDQDNGLYSL-GRFGDKGLLPQKSTIVPAGSQYENTF 501
Db 467 LHNEYSKNRPDELGLFTYNDLLCGGQYDAGCGRFQCTSTLPPKSVIVQPSGKQREHSI 526
Qy 502 YFDAALAKDIYRLNYSINTVGYRGGEYTYGYSDDDEKRAFGENSTYKHKHCNQCIGY 561
Db 527 YLDTSLQDKYQLDYSYNASQYRFSGBHASYYSQKQFQDKFGDSDQIYKQHCSPSCDVY 586
Qy 562 EPLVKYKGRKRANNSHYSISADFGDYEMPFASYSRTHERMPNIOEMYSQIGDSGVHTALK 621
Db 587 EPLVTTSGKKHAINHSVTLAKYDTGMPFVSFARTHERMPNIOEMFFSQIGDGVNTALK 646
Qy 622 PERANTWOFQNTYKGLLKQDDTLGLKLVGYSRIDNYIHNVYKWMDLNGNIPSWSS 681
Db 647 PEQANTYQLGFNVFKRNLTDNDTLGLKVVGYSRINNYIHNVYKQYDTK-NPPSWTS 705
Qy 682 TGL-AYTIOHRNFKDKYKHKHGFLELNYDYGREFTNLSYAYOKSTQPTNFSASESPNA 740
Db 706 GALKGDTIQHRNWQMPVHKOGLELEINYGARYETNLSYARQKTDQPTNYSASESPNS 765
Qy 741 SKEDQLQGYGLSVSALPRDYGRLEVYGTWLGKGLTGGAMRYFGKSIIRATAERYIDG 800
Db 766 SKEDQLTQGYGLSVSMPLPRDYGRFELGVGRGFDKLTIGSAVRYGQSPRATIEPRIDG 825
Qy 801 TNGNTSNVROLGRSKQETTLARQLIFEDFYAAYPEKKNLIFRAEVKNLFDRIYIDPL 860
Db 826 THGNTSHSDDKGAHVIKQIEMKLARQLVHDFYVAYEPIKDLVMRLDVQNAFDKLYIDPL 885
Qy 861 DAGNDAATQRYSSFDPKDDEEYTCNADKTLG--NGKYGTGTSKSVLTNEARGTFLITM 918
Db 886 DANDAATQRYHSI-YNDADEGAPCAAGQ-LCKPDAKYGGTTTSVLTNEARGSRLLFSM 943
Qy 919 SYKF 922
Db 944 TYKW 947

RESULT 9
US-09-303-518D-876
; Sequence 876, Application US/09303518D
; GENERAL INFORMATION:
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Massignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisserial Antigens
; FILE REFERENCE: CHIR0160
; CURRENT APPLICATION NUMBER: US/09/303,518D
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 876
; LENGTH: 393
; TYPE: PRP
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (21)..(21)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (104)..(104)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (163)..(163)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (165)..(165)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (171)..(178)
```

```
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (301)..(301)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (339)..(339)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (353)..(353)
; OTHER INFORMATION: Xaa= any amino acid
US-09-303-518D-876

Query Match 40.3%; Score 1976; DB 17; Length 393;
Best Local Similarity 94.1%; Pred. No. 1.7e-175;
Matches 369; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

Qy 531 GYGSDDDEKRAFGENSTYKHKHCNQCIGYEPVLKYGKRRANNSHYSISADFGDYEMP 590
Db 2 GYGSDDDEKRAFGENSTYKHKHCNQCIGYEPVLKYGKRRANNSHYSISADFGDYEMP 61
Qy 591 FASYSRTRHPNIOEMYSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKL 650
Db 62 FASYSRTRHPNIOEMYSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKL 121
Qy 651 VGYRSRIDNYIHNVYKWMDLNGNIPSWSTGLAYTIOHRNFKDKYKHKHGFLELNYDY 710
Db 122 VGYRSRIDNYIHNVYKWMDLNGDIPSWSTGLAYTIOHRNFKDKYKHKHGFLELNYDY 181
Qy 711 GRFTNLSYAYOKSTQPTNFSASESPNASKEDQLKOGYGLSVLPDYGRLVGTR 770
Db 182 GRFTNLSYAYOKSTQPTNFSASESPNASKEDQLKOGYGLSVLPDYGRLVGTR 241
Qy 771 WLGKGLTGGAMRYFGKSIIRATAERYIDGTNGGNTSNVROLGRSKIKQETTLARQPLIF 830
Db 242 WLGKGLTGGAMRYFGKSIIRATAERYIDGTNGGNTSNVROLGRSKIKQETTLARQPLIX 301
Qy 831 DFYAAYPEKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFDPKDDEEYTCNADK 890
Db 302 DFNAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAAXERYSSFDPKDDEEYTCNADK 361
Qy 891 TLCNGKYGTSKSVLTNFAARGTFLITMSYKF 922
Db 362 TLCNGKYGTSKSVLTNFAARGTFLITMSYKF 393

RESULT 10
US-09-543-681A-8096
; Sequence 8096, Application US/09543681A
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8096
; LENGTH: 760
; TYPE: PRP
; ORGANISM: Proteus mirabilis
US-09-543-681A-8096

Query Match 15.4%; Score 753.5; DB 19; Length 760;
Best Local Similarity 25.6%; Pred. No. 1.9e-60;
Matches 240; Conservative 144; Mismatches 328; Indels 225; Gaps 32;

Qy 11 CFYLMGVNLYHHS--YAEADAGRAGSEAIQVLEDVHVKAKRVPKDKVFTDARAVSTFQD 68
Db 24 CILFSGI-----HSLVFAAETKTVQFSSLVK-----SSAQKQTPQKALSPPGAYSAIGE 74
```

Qy	69	IFKSSENLDNIVRISIPGAFTQDDKSSIVSLNTRIGDSGGRVNWVDGJTQTFYSTDA	128
Db	75	I-NLSSVEQALRSTPGTYTQMDASOPGVGNTIRGLSGFRVNMMDGVTQMTYSPSQ	133
Qy	129	GRAGS--SOFGASVDSNFAGLDVVKGSFSGSAGINSLAGSANLFTLGVDVVOGNNTY	186
Db	134	YAHGGQYNGFNSMIDPNFIQIDISRGQODGENSINALAGSANFTIGDIDVLFNSHRW	193
Qy	187	GLLLGLGTNSTKGNMAAI-GARKWLESAGSVGLYGHRRSRVAQNTYRVGGGQHGN	245
Db	194	GIRSKAARGTNGLYNGWVAIAKQPIFNHDGYIGAMFALSGHNISSYKNCAG-----	247
Qy	246	FGAEYLERRKQTFVQEGGLKFNSNGKWERDP-QRPYWK-TKWYQKYNDPOELOKYIBG	303
Db	248	-----FNSHEFATTKDFNQKPHSELAKISFKPNDSEIE-----	281
Qy	304	HDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRLNTIKGSRKII	363
Db	282	-----LSGRF-----YNNFNRTTIESQYAK-----	304
Qy	364	NRNYQFYGLSLNSYANLNTAAYNSGRQYKPKSGKFTGMLGLKDPETYNNAKILDLNNT	423
Db	305	---YRYT---PLNDLFDSEILLSRSQASQKAFAGDSLMS---LRNGHAKNISNALVAKNT	354
Qy	424	ATFRLPRETELQTTLGFNYHNEYGNRRPEELGLFFDGPDOONGLYSYLGRFKGDKGLL	483
Db	355	SRENY-HELDWALTGLTKLMSIDYNR-----EVTAPSDPQWOSQNNIEY-----	397
Qy	484	POKSTIVQAGSVOYNTFYFDDAALKDVIYRLNYSNTVGYRGEGEYTYGYSDDFEKRAF	543
Db	398	---NVFAPOGKNDISLFSOMKFEYDITLALNLNTDYHKG-----	437
Qy	544	GENSPYKKHCNOCGIYEPVLKKYKGRKANNHSVS-----ISADFGDFMPFASYST	597
Db	438	---YKPCADPKAACFPE-----GAMNVNRHEKAWEPGALFSAQIIEPEFPFISYAHT	486
Qy	598	HRMPNTOEMFYSGIDSGVHTALKPERANTWQGFNTYKGLLKQDDTIGLKLGVYRSRI	657
Db	487	TRAPNPOEIFFANEGGASMPFLURSEKADTFQIGFNSYRPLDIVAGDSFLRKALMYHTKV	546
Qy	658	DNYT-----HNWY--GKWDMLNGNIP-----SWYSSTGLATYIOHRNFKDKVHKHGFLELN	707
Db	547	KNYISSDSYNVCKGGRCKIDGFTDDDYSGMEYDGNIL--YTNSLDPVTWGRGYELQVN	604
Qy	708	YDGRFPTNLISUYAQKSTOPTNFSDESASPNNAKEDQLKQG-YGLSRVSALPRDYGRLE	766
Db	605	YDADVFYTTLSYNSERTSQPTS-----LAWMDFGQSPASELPKYIYATID	648
Qy	767	VGRFWLGNK-LTLGGAMRYFGKIRATAERYIDGTNGGNTSNVRQLGKRSIKQFETLAR	825
Db	649	AGVLPDDRSLGLGTIIQTGKRKASAGDFD-----TGTIPMVEQEKI-----	694
Qy	826	QPLIFDFAAYEPKKNLIFRAEYKVLNDRRYIDPLDAGNDAATQRYYSFDFPKDKDEEVT	885
Db	695	-PTVINAYAYQMNKNVSLKFAVHNLNMKNYSNALDRSNSAPLM-----DEQ-----	740
Qy	886	CNADKTLNCKYGGTSKSVLTJNFARGTELTITMSYKF	922
Db	741	-----GONKQT-----ARGSRFLGGEIRF	760

RESULT 11

RESOLV. 11.
US-09-252-991A-28261
; Sequence 28261, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788

```

: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 28261
: LENGTH: 977
: TYPE: prt
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-28261

```

Query Match	8.6%	Score 422:	DB 15:	Length 977:
Best Local Similarity	22.6%	Pred. No. 3.3e-29:		
Matches 214:	Conservative 115:	Mismatches 382:	Indels 238:	Gaps 37:
Qy	30	RAGSEAOIQVLEHHVAKRVPKDKVFTDARAVST--RODIFKS-SENLDNIIVRSIPGA	86	
Db	211	RVSQDDLVQMSFSV-ISAR--PDDWYIQTPHSVSVIGREGIERNPPRHAADMLEETPGV	267	
Qy	87	FTQODKSGGIVSLNIRGDSGRVNTWVGITQITQYSTSTDAGRAGGSSQFPA-SVDSNF	145	
Db	268	YSSVSQODPGLSVNIRIQDYGRVNMVSDGMRQNYQ-----QSGHQORNGTLYVDPEL	320	
Qy	146	IAGLDVVKGSPSGAGINSLAGSANLRLTGLVDDVVQGNNTYG---LLLKGLTG-TNSTK-	200	
Db	321	LSEVVIDDKGASSMGAGVIGGIANFRTVEARDLVRPGKQVGRVRLTSGLGGDANGTHF	380	
Qy	201	-GNAMAAIGARKWLSEASVGLYGHSSRVSQAQNVYRGVGGQGHICFNFGAEYLRKKORYF	259	
Db	381	IGSAFAIGTEVW-----DMLVAASERHLG-DYDPTCKGS-IG-----ELRTGAWF	424	
Qy	260	VOEGGLKPNSSNGKWERDFORPYWKTWKYQYNPDQELQKYIEGHDKSWRENLAPOQDIT	319	
Db	425	NPEAGQRVK-----HSPVAYSQVY-----MSRLAKLGLVAL	455	
Qy	320	PIDFSSLKQOQSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNQVFNGLS-NSY	378	
Db	456	PDQD---RLQFSYLTQVSYDDA-NMLNTEQAOLWEKLGDSDVRAQNAIDYGVAPDNPL	511	
Qy	379	ANLNLTAAY--NSGRQYPKGSKFTGMGLKDFFTYNNAKILDLNNTATFLRLPRETELQT	436	
Db	512	VDFKAKLYYVDNRNRQOTLQRGITPGYSITYQTQTYG---AQANTSTFALDDLSTLRA	567	
Qy	437	TLGFNYFHNEYGKNR-----PPEELGLFFDQDODNGLYSYLGRFKGDKGLLPQ	485	
Db	568	NYGLEFFYDKVRPSSOPRASTSAVGPPAAEGM---TPKGDRLGLSLFARLDYD-----	618	
Qy	486	KSTIVQAGSQYFNTFYDALKKDIYRL-----NTSTNTVGYRFGGEYTYGYSDDF	539	
Db	619	-----YDDWLNLNAGLRVDRYRLRGDTGFNARTFILGTTROTMDPLQYAVD---664		
Qy	540	KRAEGENSPYKKHCNOSCGIYEPVLKKYKKRANNHUSVISADFG-DYMPFPFASYSRTH	598	
Db	665	-REEGREFSPT-----GLSVKPGVDWLQLEFATYKGW	695	
Qy	599	RMPIQIEMYSQIGDSG-----VHTALKPERANTWQGFNTYKKGLLKODDTLGLKLVG	652	
Db	696	RPPATVESLTIGRPHGGGAENWYFPLSPRSKAWEGVFNVLKENLWFSDDRGLGLKAY	755	
Qy	653	YRSRIDNYIHNVYKGWDLNGLNIPSWSSSTGLAYTIQHRNFKDKRVKHGHFELELNYYDGR	712	
Db	756	FDTRVDDFDIFMGMGQ-----PPGYGMAGIGNSAYVNNL-DSTRFRGEVYQJDYDAGL	807	
Qy	713	FTFWLNLAYOKS-----TQPTNFSDAESPNNASKEDQIKQYIGLSRV---755		
Db	808	AYGOLSYTHMIGSNDFCSTAWLGGVATQYKVGSGRRPVIDMRPDEQANAATHCSAVLGS	867	
Qy	756	SALPRDYGRLVCTRWLGNKLTLCGAAIRYFGKSIIRATAERYIDGTNGGNTSNVROLGK	814	
Db	868	AEHMPDRG-----SLTLC--MRFDRLDVGARARYSEGYSVAGGATVSQAGV	914	
Qy	815	RSIKQETLARQPLIFDYAAYEPKKNLIFRAEYKNIPLDRRYIDPLDAGNDAAATQRYSS	874	

Db 915 YPADKEY-----TVYDYGSRYSDELTLRLAMENVTDRAVLPLG----- 956
QY 875 FDPKDKDEEVTGNADKTLGKNGKYGSTSKSVLT-NFARGRTFLTMSYKF 922
Db 957 -----DWLAFTLGRGRTLGQGTLEYQF 977

RESULT 12

US-09-252-991A-27095
; Sequence 27095, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074.788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094.190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27095
; LENGTH: 992
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27095

Query Match 8.6%; Score 421; DB 16; Length 992;
Best Local Similarity 23.3%; Pred. No. 4.2e-29;
Matches 226; Conservative 129; Mismatches 370; Indels 250; Gaps 42;

QY 47 AKRVPDKKVTDAVSTRODIPKSSENLD-----NIVRSIPGAFTQODKSSGIVS 98
Db 169 AELADPOKETV-----APRSSVYLSSEIDIRFCRVSVGDLLOGIPGVQVQSDSRNGALD 223
QY 99 LNIKRGDSFGFVNTPWDLGITFTSTSDAGRAGSGSQFGASVDSNFIAGLDVVKVGSFG 158
Db 224 VNIRGIQGSRAVVRVDGAEO-----ALDVYRGYAGTQQRSDYIDPDLVSSVYVDKGPSTR 278
QY 159 SAGINSLGASANLTLGVDDVQGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESAS 218
Db 279 SGAI---GGSVEMETIGVKDLVLDGKDLGVRFETGDVNN---GVAPQHRSSASTENLSS 332
QY 219 V-----GVLYGHSRR---SVAQNYRVGGQGHGNGFAGYELERRKQRYFVQEGGLKFNSNS 271
Db 333 VPDDRGLSFGSAKSGSAFAFR---NEHL-DLVAAYAQNRQGNVF-----S 376
QY 272 GKWERDFORPYWK-----TKWY---QKYNPQELQKYIEGHDKSWENLAPQYDIT 319
Db 377 GKGDQRYRVNRYGRESSVAKYVNAAGEEVLNSSSETESYL--LKATWR--IADEHTLD 432
QY 320 -----PIDPSSLKQOSAGNLFKPLEVDGV-FNKYTAQPRDLNLTGSRKIINRY 367
Db 433 LGYRRYDGRTEIMPDSIFREGTAGIYQYPLSEVKIDITYARYLYPEN-----NPLV 485
QY 368 QFNYGSLNSVANLNTAAYNSGRQKYPKSGKFTGWGLLKDFETYNNAKI-LDLNNTATE 426
Db 486 DLSTGLWMTDAKSDMLTSLVLAQSRQAYRSRDNWT-----RQDNRRIGGDLNNVARE 536
QY 427 RL-----PRETELQTLGFN-----YFNEYGKN-----RPEE 455
Db 537 ETDFGDKFLDGGSFQVEDIQPKSVVYTTLHDINANLRLDRATQREYGLNGKLEKFEVER 596
QY 456 LGLFEDG-----PDQDNGLYSLGREFKDGKLLPKQKSTIVQPA--GSOYF-----NTFYED 504
Db 597 LTLGGGGRYSHFNKDKMGISASPRRDRMFI-----TVSRPGYIGSMFMFPDQNGQYTD 652
QY 505 AA---LKKDIYRLNYSNTVGYRFGGTYTGYGSDDEKFRAGFNSPTYKKHCNOCGIY 561
Db 653 ATDPLRNLNGIYNTNNTNPFEGIPE-----DEEGPA---NVTVHPSRVNTNVYGY 698

QY 562 EPVLKKYKCRANNHSHYSISADF---GDYFMPFASYSRTHMPNIOEMYSQIGDSGVHT 618
Db 699 N--YSKKGSSRGCGFSPAFIGINFELAPDTFV-YASYTEGLRLPLSFE---TSQGTLOQEP 752
QY 619 A--LKPERANTWQGFNTYKKGKLLKQDDTLGLKLVGYRSRSDNYIHNVYKQWDLNGNIP 676
Db 753 GKDLKPERSRWEIGASALRDSLLADGDSAAIKLAYFNNTIKNYITRYD-----P 803
QY 677 SWVSSTGLAYTIQHRNPKDKVHKHGFELNLDYGRFFTNLSYAYQKSTQPTNFSDASES 736
Db 804 GOMGLMTFSNT-----DSYRTSGLQSHYDAGRFAFADLSATYLYLKTETCDAAFAARL 856
QY 737 PNNASKEDOLKQGYLGRVSAL-----PRDYGRLEVGTGRWLNKGLTLGGAMRYEGK 787
Db 857 RAGANRQRTENTPNCPTGSGFMGSYTTQNPRLATNLTAGLRFDDQALTUGGRMTY--TS 915
QY 788 SIRATAERYIDGTNGGNTSNVROLGRRSIKQTETLARQPLIFDFYAAEYEPKKNLIFRAE 847
Db 916 GPTATADKPQVQ-----ATTPOIEYRSVQ-----LFDLFLKYKLFHEHTELNAS 959
QY 848 VKNLFDRRYIDPLDAGNDAATQRYYSFDPKDKDEEVTGNADKTLGKNGKYGSTSKSVLTN 907
Db 960 LQNLTDRYLDPL-----AQSFMP----- 978
QY 908 FARGRTFLTMSYKF 922
Db 979 -APGRTLVRVGMQAKF 992

RESULT 13

US-09-897-516-4927
; Sequence 4927, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 4927
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-4927

Query Match 8.2%; Score 402; DB 22; Length 364;
Best Local Similarity 28.4%; Pred. No. 4.8e-28;
Matches 109; Conservative 57; Mismatches 126; Indels 92; Gaps 11;

QY 74 ENLDNIVRSIPGAFTQODKSSGIVSLNIRGDSFGFVNTPWDLGITQTFYSTDAGRAGG 133
Db 39 ESMDSVLRSLPGTYTQMDTSQGTIAVIRMGSGFGRVNMVGVDSQSYFGIAPSEFAHGT 98
QY 134 S--SQFGASVDSNFIAGLDVVKVGSFSGSAGINSLAGSANLTLGVDVDDVVOGNNTYGLLLK 191
Db 99 QPYNQTCALIDSNFIIRTDVDRGOANDSDSYNALVGSANFRTIGDIDVIFPGKLGILTK 158
QY 192 GLTGTNTKGNAMAAIGAR-KWLESASVGVLYGHRSRVSQAQNYR--VGGGGQHGIFGA 248
Db 159 SAYCTNGLKNGMTAIAAGTQAFSTEGSIGAMLAISGHSIDAHYKNAMGVSSSEFGT--- 215
QY 249 EYLERRQRYFVQSGGLKFNNSNCKWERDFORPYWKTKWYKYNKNDPOELQKYIEGHDKSW 308

```
Db 216 -----DKTENR-----EPNSQLM 228
Qy 309 RENLAPQYDITPIDPSSLKQOSAGNLFKLEYDGVF--NKYTAQFRDLNLTIGSKRIINRN 366
Db 229 KINKP-----NDFHELSGRFYHNKFT-----KRHDSYD 260
Qy 367 YQFNPGLS-LNSYANLNLTAAYNSGRQYKPKGSKFTGMLLKDFETYNNAKILDLNNTAT 425
Db 261 YLKYHVTTPFSELIDTNILLGSGKGNQYFVKMSGLCKG-----ESHKNSNTIDIKNTSR 315
Qy 426 FRLPRETELQTTLGFNFHNEYGK 449
Db 316 FNY-GETDISFTLGSKLMDEYHK 338

RESULT 14
US-60-215-161-4927
; Sequence 4927, Application US/60215161
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; CURRENT APPLICATION NUMBER: US/60/215,161
; CURRENT FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 4927
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-60-215-161-4927

Query Match 8.2%; Score 402; DB 26; Length 364;
Best Local Similarity 28.4%; Pred. No. 4.8e-28;
Matches 109; Conservative 57; Mismatches 126; Indels 92; Gaps 11;

Qy 74 ENLNIVRSIPGATQODKSSGIVSLNIRGDSGFRVNTWVDGTTQTFYSTSTDAGRAGG 133
Db 39 ESMDSVLRSLPGTQMDTSQGTIAVIRGMSGFGRVNMVYDVSQSFYGIAPSEFAHGT 98
Qy 134 S--SQFGASVDSNFIAGLDVVKGSFSGAGINSLAGSANLRTLGVDDVVOGNNYTGILLK 191
Db 99 QPYNQTAGALDSNFIITDVRGQANDSDSVNALVGSANFRTIGIDVIFEGNKLGLTK 158
Qy 192 GLTGTNSTKGNMAAIGAR-KWLESASGVLYLGHSSRRSVAQNYR--VGGGGQHIGNFGA 248
Db 159 SAYGTNGLKGNMTAIGRTQAFTGEGSICAMLAISHSDAHYKNAMGVSSSEFGT--- 215
Qy 249 EYLERRKQRYFVQEGGLKFNNSGKWERDFQRYWTKTKWYQKYNQNDPOELQYIEGHDKSW 308
Db 216 -----DKTENR-----EPNSQLM 228
Qy 309 RENLAPQYDITPIDPSSLKQOSAGNLFKLEYDGVF--NKYTAQFRDLNLTIGSKRIINRN 366
Db 229 KINKP-----NDFHELSGRFYHNKFT-----KRHDSYD 260
Qy 367 YQFNPGLS-LNSYANLNLTAAYNSGRQYKPKGSKFTGMLLKDFETYNNAKILDLNNTAT 425
Db 261 YLKYHVTTPFSELIDTNILLGSGKGNQYFVKMSGLCKG-----ESHKNSNTIDIKNTSR 315
Qy 426 FRLPRETELQTTLGFNFHNEYGK 449
Db 316 FNY-GETDISFTLGSKLMDEYHK 338

RESULT 15
```

```
US-09-897-516-6415
; Sequence 6415, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 6415
; LENGTH: 937
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-6415

Query Match 7.9%; Score 386; DB 22; Length 937;
Best Local Similarity 19.9%; Pred. No. 7.2e-26;
Matches 216; Conservative 144; Mismatches 385; Indels 338; Gaps 40;

Qy 27 DAGRAGSEAOI-QVLEDVHVHAKRVPKDKVFTDARAVSTQRDIFRSESLNDNIVRSIPG 85
Db 6 DKDEAGDYDAVYDKDISNIYIGKKEIER-----YKGASPAD-VIRGAVG 47
Qy 86 AFTQODKSSGIVSLNIRGDSGFRVNTWVDGTTQTFYSTSTDAGRAGSQFGASVDSNF 145
Db 48 VYSGDARNSCALDINIRKIQGQGRIPVTIDGTEQ-----AITVGRGYNANRRNYIDPNL 102
Qy 146 IAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVOGNNYTGILLKGLTGTNSTK----- 200
Db 103 ISSIEIEKGPRLNRRVKRSGVGAISIKTLNLDVVPGETFGINTKLETSSNSVKERTPS 162
Qy 201 -----GNAMAAIGARKWLESASGVLYLGHSSRRSVAQNYRVG-GGGQHIGN 245
Db 163 LSLGQDYRVDPNFIHNGIESDPALKITPHSSKDNKLEFGKDNA-----FRVAVGTRQYFD 218
Qy 246 FGAEYLERRRKQRYFVQEGGLKFNNSGKWERDFQ-----RPYW----- 283
Db 219 LMLAYAYRHKGNFYFAGKGG-AHYDAAATEADINLMLNPKTSLDPLPFAARIYRPGNEV 277
Qy 284 -----KTKWYQKYNQNDPOELQYIEGHDKSW-----RENLAPOYDI 318
Db 278 PNTSSKMQSVLIKNTWH--FTDEQALQALAFNTRMEFGDIMPSRLASVLAKENSVPQWPL 335
Qy 319 TPIDPSSLKQOSAGNLFKLEYDGVFNKY-----TAQFRDLNLTIG-SRKLIINRYQ 368
Db 336 -----ANARQQAASLANK--WDSAAANPYIDFNMNLWTTRTISNTNTSGGYPGCVTRDYD 388
Qy 369 F--NYGLSLNSYANLNLTAAYNSGRQYKPKGSKFTGWG--LLKDFETYNNAKILDLNNTA 424
Db 389 WEGGKGRSTWIDGTTLINTAVTNAQNRR-----WGVDISNKKFELTQNLDLTLMGNFQ 439
Qy 425 TFRLPRETEL-----QTTLGFNY----- 442
Db 440 RERLGSDDDISHIDNLYFFQSPARKQORQELAFNFDWRPTSWLALGAGAKRVSYWSKD 499
Qy 443 -FNE--YGNKRPEELGLFF-----DGD-----ODNGLYSYLGR----- 475
Db 500 DFLNERRIARDNRYKPEGEIIGYKMSYWRVTWTEDEANDKRNDEKYSKLSRTEKREL 559
Qy 476 -----FKGDKGLLPQKSTIVQ-----PAGSOYF-----NTFFYDAALKK 509
Db 560 AKTKKIKDRNDRIOQRWLVVEERFEWKYNPDTGKLNKSDNPFYNGQLDMNEKVIDPISGR 619
```

Qy	510	DIYPLNYSTNVTGVRGGEYTGYYGSDDEFKRAFGENSPYKHKHCNOCGIEYEPVLKYYG	569
Db	620	EARKYEYSANSL-----SNNDKILSQDVDP-----WEPAPKR--	652
Qy	570	KKRANHVSISADFGYEMPFASYSRTHRMPNIOE--MYFQIGDGSVHTALKPERANT	627
Db	653	KSHAWPTFSAATYITDDLRYVTRYAEAVMPSIFEDTVGFCGVRENYGLGFKPERAKT	712
Qy	628	WQFGFNTYKGLLKQDDTGLGLKLVGYRSRIDNYI--HNYVKGWMDLNGNIPSWVSSTGLA	685
Db	713	IETGVYDFESQLYNABRNADIKLSYNTYVIEVDFDRDNTY-----	752
Qy	686	YTIQHRNFK--DKVHKHGELELNVDYGRFFTNLSYAYOKSTQPTNFDAS--ESPNNAS	741
Db	753	-----NFAOLDKQKLAGLELQARYDNGSFFETDGLVYNKKVCDNNNSAARMDSQNRVY	806
Qy	742	KEDQLKQGY--GLSRYSALPRDYGRLVEVTRWLGNKLTGLGAMRVFGKSIATASERYID	799
Db	807	VPECIDGGPFGGYLRTSIQPKYATNLNVGGRFLFDEKLELGSRLMYHSRA--ENKDEKWL	864
Qy	800	GTNGGNTSNVRQLGKRSIKOTETLARQPLIFDFYAAEYEPKNLIIFRAEVKNLFDRIYD	859
Db	865	G-----VLPNEYGKISNNPMRWNSVFTVDAYSQITPAISMELTGNTNTRYIDLP	916
Qy	860	LDAGNDAATQRYYSSEDPDKDEVTGNADKTLGCKGYGSGTGSVLTNFGARTTFLTMS	919
Db	917	L-----TRSMIP--APORTFKLSLT	934
Qy	920	YKF 922	
Db	935	SQF 937	

Search completed: July 24, 2002, 08:58:25
Job time: 441 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:54:04 ; Search time 54.24 Seconds
(without alignments)
2276.355 Million cell updates/sec

Title: US-09-762-926-4

Perfect score: 4904

Sequence: 1 MRSSRLPAPICFLMGVLMY.....SVLTNFAFGRTFLITMSYKF 922

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 410475 seqs, 133914855 residues

Total number of hits satisfying chosen parameters: 410475

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA.New.*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pap.*
- 2: /cgn2_6/ptodata/1/paa/US05_NEW_COMB.pap.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pap.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pap.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pap.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pap.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	279	5.7	917	5	US-09-584-501A-11
2	269.5	5.5	908	5	US-09-584-501A-12
3	239.5	4.9	934	5	US-09-540-209B-7679
4	203	4.1	564	5	US-09-545-199F-30
5	199.5	4.1	1083	5	US-09-540-209B-9521
6	195.5	4.0	925	5	US-09-540-209B-6249
7	189	3.9	967	5	US-09-545-199F-32
8	184.5	3.8	1094	5	US-09-540-209B-8695
9	180.5	3.7	707	5	US-09-540-209B-10058
10	179.5	3.7	659	5	US-09-545-199F-105
11	177	3.6	723	5	US-09-540-209B-6565
12	177	3.6	833	5	US-09-540-209B-7143
13	172	3.5	702	5	US-09-540-209B-8474
14	172	3.5	783	5	US-09-540-209B-9746
15	166.5	3.4	930	5	US-09-584-501A-2
16	160.5	3.3	3241	1	PCT-US01-13240-1
17	160	3.3	1178	5	US-09-540-209B-7922
18	159.5	3.3	1043	5	US-09-540-209B-9606
19	159	3.2	760	5	US-09-540-209B-6609
20	159	3.2	1152	5	US-09-540-209B-10027
21	157.5	3.2	722	5	US-09-540-209B-9749
22	156.5	3.2	947	5	US-09-540-209B-9753
23	156.5	3.2	1009	5	US-09-540-209B-10198
24	155	3.2	1102	5	US-09-540-209B-9537
25	154.5	3.2	699	5	US-09-540-209B-6132
26	154.5	3.2	1111	5	US-09-540-209B-9178

ALIGNMENTS

RESULT 1

US-09-584-501A-11

; Sequence 11, Application US/09584501A

; GENERAL INFORMATION:

; APPLICANT: Lo, Reggie Y.C.

; APPLICANT: Schryvers, Anthony B.

; APPLICANT: Potter, Andrew A.

; TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS OF

; FILE REFERENCE: PASTEURELLA HAEMOLYTICA AND VACCINES CONTAINING THE SAME

; CURRENT FILING DATE: 1996-11-29

; CURRENT FILING DATE: 1996-11-29

; PRIOR APPLICATION NUMBER: 08/753,759

; PRIOR FILING DATE: 1996-11-29

; PRIOR APPLICATION NUMBER: CA 2,164,274

; PRIOR FILING DATE: 1995-12-01

; PRIOR APPLICATION NUMBER: 60/008,569

; PRIOR FILING DATE: 1995-12-01

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 917

; TYPE: PRT

; ORGANISM: Neisseria gonorrhoeae

US-09-584-501A-11

Query Match

Best Local Similarity 5.7%; Score 279; DB 5; Length 917;

Matches 224; Conservative 127; Mismatches 323; Indels 376; Gaps 58;

```

QY 5 FRLAPICFLMGVLMYHHSVAEDAGRAGSEAOIQVLEDVHVYKAK--RVPDKKVFDTARA 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 FRLNLCLSLMTAL---PAYAENV-QAG-QAQEKOLDFTIQVAKKQKTRRDNVEVTLGLKL 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 63 VSTRQDIFKSSG-NLDNIVRSIPG-AFTQDK--SSGIVSLNIRGDSGFRVNTWVDGIT 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 VKTADTLTKEQVLDIRDLTRYDPCIAVVEQGRGASSG---YSIRG-WDKNRVSLTVGLA 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 119 Q-TFYSTSTDAG--RAGSQSFGASVDSNFTAGLDVVKVSGSGSAGINSLAGSNLRTLG 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKSGNSVEQSGALAGSVAFOFKT 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 176 VDDVQQNNYGLLLK-GLTGTNS--TKGNAMAA-IGARKWL--ESGASVGLVGHSSRS 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 ADDVIGEGRWGIQSKTAYSGKNRGLTQSIAGLAGIGGAELIIRTRHAGEIRAH---- 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 230 VAQNYRVGGGQHGICNGCAEYLERKQYFVQEGGLFNSNGKWERDFQRPYKTKWYQ 289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
Db 234 -----EAGRGVQSFNRLAPVDDGSKYAFIVEBECK-----NGHECKAMP----- 276
Qy 290 KYNDPQELQKIEGHDK-----SWRENLAPOVDITPIDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KQDVGDKRQTVSTRDYTCNPFLEA--DPLSYESRWLFRPGFRFENKRHY 326
Qy 340 DGVFNKYTAQ---FRDLNWKIGSRKII-----NRYQFNNGYGLSLNSYANL 381
Db 327 IGGILERTQQTDFTRDMTPAFLTKAVFDANQKQAGSLRGNKHKAYAGNH-----KYGGL 381
Qy 382 NLTRAAYNSGRQKYPKSGKFTGWLLKDFETYNNAKILDNLNTATFRPRETELOTTIGFN 441
Db 382 -----FTSGENNAAPVGAIEY-GTGVFYD-----ETHTKSYGLE 413
Qy 442 YFNEYGKNRPEELGLFFD--GPDQDNGL-----YSVLGRFKG 478
Db 414 YVYNADKTDWADYARLSYDRQGLGLDNHFOQTHCSADGSKYCRPSADRFPSY---YKS 470
Qy 479 DKGLLPQKSTIVQAGSQYFNFTFYDAALKDIIYRLNYSNTVGY-RFGG-----EYTG 532
Db 471 DRVIYGESHKLQAAFKKSEPTAKI-----RHNLSVN-LGYDRFGSNLRHQDY--Y 518
Qy 533 YGSDDEPKRAFGENSPYKKHCNOSCGIYEPVLKYYKKRA-----NNHSVTSAD--- 583
Db 519 YOSAN---RAYSLKTPP-----QNNGKTSPPNGREKNPYWVSGRNVV 559
Qy 584 -----FGD-----YEMPEA-----SYSRTH----- 598
Db 560 TRQICLFGNNTYTCTPRSNGKSYAAVRDNVRLGADVGAGLRDYDRSTHSDGSGVS 619
Qy 599 -----RMPNIQEMFYSGIGSGVHTALKPERANT 627
Db 620 TGTHTLSWAGIVLKPADWLDLTYRTSTGTFRLPSEAFEMYGWRSQDKIKAVKIDPEKSFN 679
Qy 628 WQFG-----FNTYKGLLKQDDTLGLKLVGRSRIDN-----YIH- 662
Db 680 KEAGIVFGDPGNLEASFNFNAYRDLIVR-----GYEAQIKDQKQVKGPNAYLNA 730
Qy 663 -----NVYK--W---WDLNGNIP--SWVSSSTGLAYT-IQHRNFKDKVHKHGFLELN 707
Db 731 QSARITGINILGKIDWGVND---KLEPGWYST--FAYNRVRVDIKRADRTDIQSHL- 784
Qy 708 YDGRFFTNLSYAVOKSTQPTNFSDESPPNNAKEDQLKOGYGLSRVSAALPRDYGRLEY 767
Db 785 FD-----AIQPSRYV-----VSGSYDQPECKWGVNGMLT-----YSRAKE 819
Qy 768 GTRWLGKLTILGGAMRVFGKSIRATAERYIDCTNGTNSVRQLGKRSIKQETILARQP 827
Db 820 ITELGRALLNG-----NSRNTKATARRTRPW 847
Qy 828 LIFDFYAAEPKKNLIFRAEVKNLFDERRYI 857
Db 848 YIVDVSGYTVYKHFHTLRAGYNNLLNRYV 877

RESULT 2
US-09-584-501A-12
; Sequence 12, Application US/09584501A
; GENERAL INFORMATION:
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Schryvers, Anthony B.
; TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS OF
; FILE OF INVENTION: PASTEURIELIA HAEMOLYTICA AND VACCINES CONTAINING THE SAME
; FILE REFERENCE: A34763 021645.0106
; CURRENT APPLICATION NUMBER: US/09/584,501A
; CURRENT FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: 08/753,759
; PRIOR FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: CA 2,164,274
; PRIOR FILING DATE: 1995-12-01
; PRIOR APPLICATION NUMBER: 60/008,569
; PRIOR FILING DATE: 1995-12-01
```

```
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-584-501A-12
```

```
Query Match 5.5%; Score 269.5; DB 5; Length 908;
Best Local Similarity 21.0%; Pred. No. 1.1e-12;
Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;
```

```
Qy 5 FRLKPICFYILGMVLYHHSYADAGRAGSAQIOVLDDVHVKAK--RVPDKKGVFTDARA 62
Db 7 FRNLICLSMTAL---PVYAENV--QAEQAEKQLODTIQVKAQKQTRDRDNEVTGLGKL 61
Qy 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRGVNTMVDGIT 118
Db 62 VKSSDTLSKEQVLNIRDLTRYDPCIYAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGVS 117
Qy 119 Q-TFYSTSTDAG--RAGSSQFQASVDSNFIAGLDVYKGSFSGSAGINSLAGSANLRTLG 175
Db 118 QIOSYTAQAALGCTRTAGSSGAINIEIYENYKAVEISKGSNSSEYGNALAGSAVAFQTK 177
Qy 176 VDDVVOGNNTYGLLLK-GLTGTNSTKGNMAAIGARKWLESASGVLYGHRSRSVAQNY 234
Db 178 AADIIGEGKQWGTQSKTAYSGKDHALTQSIALAGR---SGGAELLIIYTKRRGRETHAH 233
Qy 235 RVGGGGQHIGNFGAAYLERRKQ-----RYFVQV---GGLKFNNSN-----SGKWDERF 278
Db 234 KDAGK--VQSFNRLVLDEDKKEGGSQYRYFIVEECHNGYAAACKNKLKEDASVKDER-- 289
Qy 279 ORPYWTKTKYQKNDPOE-LQYIEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKL 337
Db 290 -----KTVSTQDYTGSRNLLANPLEYGSQSWL--FRPGW-----HLDNR 326
Qy 338 EYDGVFNKYTAQFRDLNLTGSRKRIINRNYQFNYSLSNYSANLNTAAYNSGRQKYPK- 396
Db 327 HYGVAVLERQQOTFDTRDMIVPAFYFSEDYVPGSLKGLGY-----SGDKAERL 376
Qy 397 -----GSKFTGWGLKDDFETYNNAKILDLNNTATFRPRETELOTTLGFNYFHNEYKGR 451
Db 377 FVQEGESTLOGIG-----YGTGVFYD-----ERHTKNRYGVYVYVHNADKDT 418
Qy 452 FPEELGLFFD--GPDQDNGL-----YSVLGRFGDKGLLPKST 488
Db 419 WADYARLSYDRQGLDLNRLQOQTHCSHDGSKNCRPDGNKPYSF---YKSDRMIIYESRN 475
Qy 489 IVQAPAGSQYENT-----FYFDAAL----- 507
Db 476 LFOAVFKKAFDTAKIRHNLSINILGYDRFKSQLSHSDYIYLQNAVQAYDLITPKKPPFNGS 535
Qy 508 KKDIYRLNYSNTVVG-----YRFGGE-YT-----GYGSDDEFKRAFGENSPYKK 552
Db 536 KNPYRVISGKTTVNTSPICRFGNNTYDCTPRNIGNGYI-----AAVDNVRGLR 587
Qy 553 HCNOSCGI-YEPVLKKGKRRNNHVS-----ISADFGYFMPFASYSRTH-----R 599
Db 588 WADVAGAIRYD-----YRSTHSEDKSVSTGTHRNLNWSNAGVVLKPFMTWDLTYRASTGFR 642
Qy 600 MPNIQEMFYSGIGSGVHTALKPERANTWQFG-----FNTYKGLLKQDDT 645
Db 643 LPSFAEMYGWRACESLTKLTLKPEKSFNREAGIVFKGDFGNLEASYFNNAIRDLI----- 697
Qy 646 LGKLKLVGRSRIDN-----YTH-----NVYK--WVDLNGNIPSWSVSTGL 684
Db 698 -----AFGYETRTQNGTTSASGDPGYRNAQARIAGINILGIDHWGVMGLPGLXST-L 752
Qy 685 AY-----TIQHRNFKDKVHKHGFLELNIDY--GREFTNLSYAYOKSQPT 728
Db 753 AYNRKVKQADIRADRTFTVTSYLFDAVQPSRYVLGICYDHPDGIWINTWFTYSK----- 807
```

Db	531	AGSFF-----VNOKLKVGDIVYR-----	54
QY	581	SADFGDYFMPFASYSR--THRPNPIQEMYFSQIGSDGVHTALKPERANTWQGFENTYKKG	639
Db	549	--DFDGYVMSEGVFAQGEYNRDKLSAFISGSVSNTGY-----WYDRFYDYDKAH	595
QY	640	LKQDDTLGLKLVGRSID-NY-----IHNVYGKWDLNGNPISWVSSTGLAYTIOHRNFK	694
Db	596	AKSKT---VNFIGNAKGGUNLYNLTEHNHNVFANIGYIS-RAPFFSGGAFLNSTVSNATNP	651
QY	695	DKYHKHGFLELNDVYGRFFTNLSY-AYQ-----KSTQPTNFSDASES--PNNASKE	743
Db	652	DANNEKVFEISGYGYRSSPLTVNINAYHTRWMDKTTTRSQDITNYESGLSEPYDASKL	711
QY	744	DQLKQYGCLSRVSALPR-----DY-----GRLEVGT--RWLGNKLTGLGAMRYFK	787
Db	712	VSTKSVINMQGVNALHOGVELDFVAKPQFQWDLSGMFSIGNWRWDSN---ASGFTVEG	768
QY	788	SIRATAERYIDGTN-----GNTSNVRQLG-----	813
Db	769	FVN--SASIKSGDGKDVTVLVNAAANGLEPOTMKNLKDVKVGGSAOTTAALGATFFDKA	827
QY	814	-----KRSIKQTETLARQPL--JFDVYAAYE---PKKNLI	843
Db	828	LRLGIDWNLYARNYADWSLNSNDLVNMSKDFSTPWRIPASTFDLNASYKFNFGKLN	887
QY	844	FRAEVKNLFORRYIDPLDAGNDAATORYYSSF	875
Db	888	LSGVNVLFPDQTYISDATDGSNHDWKTAYNVF	919
RESULT 4			
US-09-545-199F-30			
; Sequence 30, Application US/09545199F			
; GENERAL INFORMATION:			
; APPLICANT: Lowery E., David			
; APPLICANT: Fuller E., Troy			
; APPLICANT: Kennedy J., Michael			
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions			
; FILE REFERENCE: 28341/6227.NCP			
; CURRENT APPLICATION NUMBER: US/09/545,199F			
; CURRENT FILING DATE: 2000-04-06			
; PRIOR APPLICATION NUMBER: 60/153,453			
; PRIOR FILING DATE: 1999-09-10			
; PRIOR APPLICATION NUMBER: 60/128,689			
; PRIOR FILING DATE: 1999-04-09			
; NUMBER OF SEQ ID NOS: 165			
; SOFTWARE: Patentin Ver. 2.0			
; SEQ ID NO 30			
; LENGTH: 564			
; TYPE: PRT			
; ORGANISM: Pasteurella multocida			
US-09-545-199F-30			
Query Match 4.1%; Score 203; DB 5; Length 564;			
Best Local Similarity 20.2%; Pred. No. 1e-07;			
Matches 132; Conservative 84; Mismatches 273; Indels 166; Gaps			
QY	41	EDVHVKAARVPDKKKVTFDARAVSTRQDIPKSSENLDNIVRTPGAFTQODKSGGIVSLN	100
Db	44	EDNKLPGRSVLKQKNIDQ-----QQADNAADLINILPGVMAGGFRPGGQTLN	91
QY	101	IRGDSGGRYNTVMVGTITQTFYSTSTDAGRAGGSGQFASVDNSTAGLDVVKSGFSGSA	160
Db	92	INGMDAEDVRVQLDGATKSFKE-----YQGSIFIEELURKKVTVDKNGTSPQY	141
QY	161	GINSLAGSANTLGTGDDVVQGNNTYGLLLKGLTGTNS-----TKGNMAAIGARKWLES	215
Db	142	GNGCFAGTVVFETKDATDFLKENQIGGLFK--YGNNSNNQKTYSTALVLQNEKNIDL	199
QY	216	GASVGVLYGHSHRRSVAQNYR-----VGGGGQHIG-----NFGAETLERRKORYFVQ	261

Db 200 -----LLFGSVRN--ASNYTRPDKSKILFSKNKQSGLIKVMQITPEHLTLTSSVYGIH 252
QY 262 EGGLKFNSSGKWERDQRPYWKTKWKYKNDPOELQYIEGHDKSRENLAPQYDITPI 321
Db 253 KG-----WEP-----WAAK-RDVMRSRPTETETEKHYGIDVAMKRLV-----287
QY 322 DPSSLKQOAGNLFKLEBYGVFNKYTAQFRLNLTGKSRKLIINRYQFNYGLSLNSYANL 381
Db 288 -----YRD-----QKDESLSKYRYLPENKKNILN 312
QY 382 NLTAAYNSGROKYPKSGKFTG--WGLL--KDFETYNNAKILDLNNTATFRLPRETELOTT 437
Db 313 SVQLSYSKTEONDRHREKVTSSFLGTGKNSWITSOL--TFDISNTSTLNIGR-AEHELL 370
QY 438 LGFNYFNEYKGRFPELGLFDPDODNGLYSYLRFRGDKLLPKQSTIVQAGSOY 497
Db 371 FGLQWLKANK--RNTLMYHKG-----GVKKADYNYGYFQPY-----YMPSGRQY 411
QY 498 FNTFYDAALKKDIYRLNYSTNTVGYRFGGEYTYGYSDDDEFKRAFGEN-SPTYKKHCNQ 556
Db 412 TOAFVLQDQIKWQNFLE-----TGIRY-----DHINNIGOKNLAPRYN--DI 452
QY 557 SCGIYEPVLKYGKRRANNIS--VSIADFGDYMPFASYSRTHRMNIOEMY-----FSQI 611
Db 453 SAG-----HDYSQKNTNGWSYILGLKYDYNHYLSLFTNFSKTRWRAVIDEYETOYSQA 506
QY 612 GDSGVHTALPERANTWQGFNTYKKGKLLKODDPLGLKLVYRSRIDNTHNVYG 666
Db 507 SVSATSINLEKEMINQTRVGGIITLNLHFOENDAFQFRTTYFYNRGKNEIFKTRG 561

RESULT 5
US-09-540-209B-9521
; Sequence 9521, Application us/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR
; FILE REFERENCE: 2709, 1001-001
; CURRENT APPLICATION NUMBER: US/09/540, 209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 9521
; LENGTH: 1083
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-9521

Query Match 4.1%; Score 199.5; DB 5; Length 1083;
Best Local Similarity 20.4%; Pred. No. 5.1e-07;
Matches 199; Conservative 132; Mismatches 367; Indels 277; Gaps 57;
QY 4 SFLRKPICFYLGMVLMVHSHSVAEDAGRAGSEAOIOLV--EDVHVKAKEV---PKDKKV- 56
Db 86 SLKVAPNSQELSVYI-----GKQOTIKVSESTYNIYVLEKDAEYLDVGVVGVGSKKVN 141
QY 57 -----FTDARAVSTRODIFKSSENLNIIVRSIPGAPTQODKS-----SGIVSLNIRGDSG 106
Db 142 VTGAVGMVSAEALEAR-PVANASQALQGVV---PGLNLTVGNNGGALDGLTNMNIIRAGT 197
QY 107 FGRVN-----TWVDGITQTFYSTD-----AGRAGSSQFGASVDSNFIAGLDVVKG 154
Db 198 IGDGSSPLVLIDGIEGLDNTVPNDIESVVLKDAASAIYGARASFGVI--LVTTKS 255
QY 155 SFGSSAGINSLAGSANLR-TLGVDDVVQGNNTYGLLLKGLTGTNSTKGN-----AMAA 206
Db 256 GKSGKTNV-SYSGSARFSAIDGVPDIM---DSY-TFAQYFNFRASANKGGGDFAPAVMER 310
QY 207 IGA-----RKWLESASVG---VLYGHSRRSVAQNYRVGGG---GOH 242
Db 311 IKAYQETLAKATTVDNAGAGIYWQKWANANGDTDWEEFYDHWAPSQEHNLSINGTDTQY 370

QY 243 IGNGFAEYLERRKORYFVQEGGLKFN-----SNSGKWER-DFORPYW 283
Db 371 L--ISSFLD---QKGLMRHGKDKFORYTLNGKITTAFTWDFKVTYTKWTREDFERP-- 423
QY 284 KTKWYQKYNQPOELQKYIEG---HDKSWRENLAPOYDIT--PIDPSSLKQOOSAGNLFKLE 338
Db 424 -----SYLTGNFFHNLARKWPVHPAYDPNGFPMDGEVEQEMENGKQNSQ 468
QY 339 YDGVFNKYTAQFRLD-NTKIGSRKLIINRYQFNYGLSLNSYANLNLTAAYNSGRQKYPKG 397
Db 469 KDFYTNQLOLVFEPFKNMKNLNGSVERTTQYQHWELVPY-----AYNAVGDYV--- 518
QY 398 SKFTGWLKLDKFETYNNAKILDLNNTATFRLPRETELOTTLGFNYFNEYKGRFPEELG 457
Db 519 --TYVWDMGYSVAAGSSRVNE-----YSWKENYTYTNIYSDFK 556
QY 458 LFEDGPDQDNGLY-SYLGFRFGD---KGLLPQKSTIVQAGSOYFNTF-----YF 503
Db 557 QF-----DNHGVFKVMAGFNAELYKTRNITAENKTLITP-GVPTINTATDDPOAYGGYA 609
QY 504 DAALKKDIYRLNYSTNTVGYRFGGEYTYGYSDDDEF--KRAFGENSPTYKKHCNOSCGIY 561
Db 610 DNSVAGFFARVNWSYKD---RYMFEANGRYDGSRSRFGKERWG-FFPSFSAGWNIA---R 662
QY 562 EPLVKYKGRKRRANNHSVISISADFG-----DYFMPFASYSRTHRMNIOEMYFSQIGD 613
Db 663 EPEMESFAER-INMGSLKLRSACQLGNTNTDAWYFF-----YQTMP-VGSNYSGLVLNG 715
QY 614 SGVHTALPKP-----ERANTWQFG-----FNTYKKGKLLKQDDTLGLKLVYRSRIDN 659
Db 716 ERPNYATNPGISVSSKKTWETVETWVGLDWSFFNRLSG-----SFDY 758
QY 660 YIHNVYKGMWDLNAGNIPSWSSYGLAVTIOHRNFKDKVHKHGPPELNY-----DYGR 712
Db 759 FVRYTY---DMIGPAPELSLLGT--SVPKINNSD-MKSYGGELEVNWRDRICEVSYGA 811
QY 713 FFTNLSYAYQKSTQPTNFDASPNNAKEDQLKQCYGLSRV-----SALPRDYG 763
Db 812 KKV--LSDDQOKILR---YPNDSYDVGVYKGEHLNDIWLTTIGIAKSQEBMDAHLAKVD 867
QY 764 RLEVGVTRW-LGNKL--TLGGAMRYFGKSIRATAEERVIDCTNG-GNTSNVROLGKRS--I 817
Db 868 QSSVGTNMGVGDIMYADLDG-----DGKISNGTNKLGDTGDIYIIGNSPRF 914
QY 818 KQETELAROPLIEDF 832
Db 915 KYGITLDAAMKGFDF 929

RESULT 6
US-09-540-209B-6249
; Sequence 6249, Application us/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES
; FILE REFERENCE: 2709, 1001-001
; CURRENT APPLICATION NUMBER: US/09/540, 209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6249
; LENGTH: 925
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6249

Query Match 4.0%; Score 195.5; DB 5; Length 925;
Best Local Similarity 20.0%; Pred. No. 8.4e-07;
Matches 156; Conservative 104; Mismatches 257; Indels 263; Gaps 41;
QY 34 EAQIQVLEDVHVAKRVPKDKKVFETDARAVSTRO-----DIFKSSENLDNIIVRS 82

```
Db 236 EETHMLDEVITISGRIONVKSTOLGAETLRPTOLKNIPMALGEVDLLK-----MVQA 288
Qy 83 IPGAFTQODKSSGIVSLNIRDSGFGVRVNTMVDGITQTFYSTS-----T 126
Db 289 LPGVKTVGEASSG---FNVRGATQDNLILLNDG---TIYPNHLFGFFAAFNDSMVKEA 342
Qy 127 DAGRAGSSGOFASVDSNF-IAGLDVVKGFSFSAGINSLAGSANLRTLGVDD-----VVQ 181
Db 343 EYKSSIPAGGRISILDITKEANKKEFTGAGIGLVTSKLNLEIPIIKORTSVLLS 402
Qy 182 GNNTYG-LLLKGLTGNSTK-GNA-----MAAIGARKWLES-GASVGVLVYGHRSRRAQNY 234
Db 403 GRTTSDWIMKQPEKSEYKNGTAGFYDLAAIVAHFENDKHSNLVYGYSHDRFAFNSNE 462
Qy 235 RVGGGGHIGNFAGAEYLERKQRYFVQEGGLKFNSNGKWERDFQRPYWKTKYQYNDP 294
Db 463 KYG-----YNNLAA-----SARWAFVN-----480
Qy 295 QELQKYTE-CHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDL 353
Db 481 EKLIGFSAGYD-----HYD-----NNRETWNASTAYKLSFD--INQYF-----518
Qy 354 NTKIGSRKIIRNYQFNYGLSLNSYANLNLTAAYNSGRQYKPGSKFTGWGLLKDFETYN 413
Db 519 -VRADFTNILLADKHTLNFPGKSMLY-----HINSGTYE-PEGS-----ESFV 558
Qy 414 NAKILLNNT--ATFRLPRETELQTLIGFNYFHNEYKGNRPBEELGFFDGPDDNGLYS 471
Db 559 KDVLOKDALETAFAFLGDWEITPKLSVN-----AGTY-----SLFS 597
Qy 472 YLG---REFGDKGLLPQKSTI---VQPSQOYFNTFYDFAALKKDIYRLNYSNTVGYRF 525
Db 598 ALGPRSYQVAGMLPHESTITDTITAGAGKFMKTYH-----634
Qy 526 GGBY--TGYGSGDEKRAFGENS-----PTYK-----551
Db 635 GPEERLSARVAFYDNFSVAKAGFNSMRQYIHKLSNTVIMSDPTDWKLSVDVNIKQORGWQAA 694
Qy 552 --KHCNOCGIOYEPVLVKYKKKRRNNHVSISADFGDYFMPFASYSKTRHMPNIQEMYS 609
Db 695 AGLYLSNPSGSIWEYSVBGYKRR-----MSDYLDY-RGGAKLLMHHIET--DVINT 742
Qy 610 QIGDSGVHTALKEP--RANTWQFGFNTYKKGKLLKQDDTLGLKLVGYSRIDNYIHNVYK 667
Db 743 QGHAYGVELQVKQVKGKNGWM--SYTYSRTFLRQND-----KRIEKPVN--GD 788
Qy 668 WBDLNGNIPSWSSSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFT--NLSYAYOKSTQ 726
Db 789 WYPTDYKPHDFKFEVG-----NYK-FTHRYSMSINVDYSTGRPTTIPAGQYDESTQ 839

RESULT 7
US-09-545-199F-32
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David
; APPLICANT: Kennedy J., Troy
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/6227.NCP
; CURRENT APPLICATION NUMBER: US/09/545,199F
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-09-545-199F-32
```

```
Query Match 3.9%; Score 189; DB 5; Length 967;
Best Local Similarity 19.2%; Pred. No. 2.9e-06;
Matches 210; Conservative 125; Mismatches 337; Indels 424; Gaps 54;

Qy 39 VLEDVHVHAKRVPKDKVFTDARAVSTRQDIFKSSE-NLDNIYRSIFGARTQODKSSGIV 97
Db 22 ILADSHQEAPEL-----DTITVSSQDDEMNIEKKIGETVKTASOLAKRQOVDSDRL 73
Qy 98 SLAIRGDSGFGVRVNTMVDGITQTFYSTSDAGRAGSSGOFASVDSNFIA-----GLDWW- 152
Db 74 --VRYETGV-----TVVEAGRFGSSGYAIRGVNDENRVAITVDGLHQAE 114
Qy 153 -----KGSFSSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGNSTKG----- 201
Db 115 TLSQOGKELFEYGNFNTRNSVEIETLKVAKIAKADSVK-----GSGSLGGAFLV 168
Qy 202 --NAMAAGARKW--LESGASGVLYGHSRVSVAQNYRVGGGQHGIGFAGAEYLERK 255
Db 169 ETKDARDLFTEKDMHIGYKAGYSTADNOGLNAVTLAGRYQM-----FDALIMHSKR 219
Qy 256 QRYFVQEGGLKFNSNGKWERDFQRPYWKTK--WYQKYN-DPOELQKYI-----E 302
Db 220 HGHELENYDYKNGRDIQGEREKADPYITIKESTLVKFSPSTENHRFTVASDITYLQHSR 279
Qy 303 GHDKSW-----RENLA---POYDITPI-----321
Db 280 GHDLSYNLVATHTQLDEKESRHANDLTRKNVSFTYENTVTVPFMDTLKLSYSQORITT 339
Qy 322 -----DPSSLKQOS-----AGNLFKLEYDGVFNKYTAQFRDLN---354
Db 340 RARTEDYCDGNELCDSYKNPLQFQKDGQILDPAKNKIKQGGSL-----STQIVDENGKP 395
Qy 355 --TKIGSR-----KIIRNYQFN-YGLSLNSY-----378
Db 396 FPPTTGTNNAAFSNNLRLRPTGFWLDCSVFDCNKPFVYVINSNCTYQAREVILLSEETVD 455
Qy 379 ANLNLTAAYNSGRQKY---PKGSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRETELQ 435
Db 456 GKLYTAKKEGGLPNYLILPNSKGYLPYD-YKERDLNTNTKQINLDTKTF-----505
Qy 436 TTLGFNYPHN-EYG--KNRFEPEL--GLFEDG-----PDQDNGLYSLG--RFGKDGK- 481
Db 506 --LTFNLENLSYGGVYSRIEKEKINKAGYEGRNPTWADRILGQSSYCYGNALCKPKHE 563
Qy 482 -----LLOPQKSTIVOPAGSOYFNTFYDFAALKKDIYRLNYSNTVGYR-----FGGEYT- 530
Db 564 PLTSFLIPVEATP-----QSLYFANILKVH-----NMISIDLGYRHDHIKYNPEYTP 610
Qy 531 -----GYI-----GSDDEFKRA-FGENSPYTKKHCNOCGIOYEPVLKKYK 569
Db 611 GVTPKIPDDMVKGLFIPMKPEPOLKDFDYNVAKFGE---AYKK-----WKLEYL 655
Qy 570 KRA-----NHSVSIADFG--DYFMPFASYSRTHRMPNIQEMYS-QIGD 613
Db 656 PKNAEENIAYIAQDKTFKHSYSLGATFDPLNLRVQVKYSKGRFAPTSDELYFTFKHPD 715
Qy 614 SGV--HTALKPERANTWQFGFNTFYDFAALKKQDDTLGLKLVGYSRIDNYIHNVYKQWDL 671
Db 716 FTILPNVLAPEEAKNOEIALTVHDNMGFVSTSVFQTK---YRHFID-----759
Qy 672 NGNIPSWSSSTGLAYTI---QHRNFK-----DKVHKHGFLELNYDYGRFF-----714
Db 760 -----LAYLSRNLNSVGGQAQARDQVYQNVNVDNAKVGLKLEINARLNLYGFHVLDFG 815
Qy 715 -TNLSYAYOKSTOPTNPSDASESPNNA-----SKEDQLKQGYGLSRVS-----756
Db 816 NTSYKFTYQGRILD-----GDRPMNAIQPKASVFEGLGYDHKENKFGADLYITRVSEKKA 869
Qy 757 -----ALPRDYGRLEVGTWRWGLNKLTLGGAMRYFGKIRATAERYIDGTNGTNSVRQ 811
Db 870 KDTYNMFYKEQGYKDSAVRWRSDDYTLVDVAVGI-----903
```

QY 812 LGRSIRKOTETLAROPILDFYAAEPKKNLIFRAEYKNLFDRRY-----IDPLDAG 863
Db 904 -----KPIKNTLQFGVYNLTDRKYLWTWESARSIKPFGTS 938
QY 864 N-----DAATQRYYS 873
Db 939 NLINQKTGAGINREYS 954
RESULT 8
US-09-540-209B-8695
; Sequence 8695, Application us/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540, 209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 8695
; LENGTH: 1094
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-8695

Query Match 3.8%; Score 184.5; DB 5; Length 1094;
Best Local Similarity 20.1%; Pred. No. 8e-06;
Matches 188; Conservative 132; Mismatches 312; Indels 303; Gaps 50;
QY 5 FRKPCIFCLMG-VMLYHSHVAEDAGRAGSEAOIQVLEDDVHVKAKRVPKDKVFTDARAV 63
Db 9 YRLKSIFILFSCIFILQTAFAQNNVITIKKNITLQEALREVEKQSDYLIAFNESKLE 68
QY 64 ST-RQDIFKSSNDLVRSIPGAF-----TQDSSSGIVSLNIR 102
Db 69 KTRVNLNINAESLDKTLASITGLSYKIDKIKYIMIPQSKVEVESKRLSGIVK-DDK 127
QY 103 GDSGFRVN-----TWVDG-----TQTFYSTSTSDAGRAGSSQF--- 137
Db 128 GDPLIG-VNVSFKGSPGTGIDGRFSILAAGNIIEFVGTQYIIVGDASSLTVV 186
QY 138 ---GASVDNFIAGLDVVRKSGSAGINSLAGSANLRTLGVDVVGQNNTYGLLLK-- 191
Db 187 LEEDAKALDEVVYVYALGKRAEKALSYVQVKSDA-----INDVKDANFVNGLTCKVA 240
QY 192 GLTGTNTKGNAMAAIGARKWLESAGS--VG---VLY-----GHSRKSVAQNYRVGG 238
Db 241 GVSINRSSGIG---GATRVVWRGAKSIVGNVNNVLYVWDGMPICNPKEINNXYSTPG 296
QY 239 GGOHIGFAGAYLE-----RRQRYFVOEGGLKFNNSNGK 273
Db 297 GGEGISDFNEDIESILGPPAAALYGSAAANGVILINTKKG---OEGKLKISIN-- 351
QY 274 WERDFQRYWKTQYQKYNQPOELQKIEGHDKSWRENLAPOYDITPIDPSSLQK---Q 329
Db 352 -NTEFMPYVMPFQNYGN-----AKGYSKSWGEMLO-----QPSTPRKDFFK 395
QY 330 SAGNLFKLEYDGVFNKYTAQFDRLNTKIGSRKII--NRNQFYNGL-----SLNSVANLN 382
Db 396 TGANIMNAANFSVGNKNQOTFVSAT--TNSTGIIPNNEYRYNFTLRNTASMLNDKLHLD 454
QY 383 LTAAY-----NSGR-----QKYPKSGKFTGWLKDFEYVNNAKIILDNLTAT 425
Db 455 LGASYVLQGGONMLSAGRYFNPLVPLFPRGEDFEA---VKYERY-----DTRN--- 502
QY 426 FRLPRETELQTLGTFNHFNEY-----GKNRFPPEELGLFFQDGDODNGLYSL-- 473
Db 503 -KPIQEWISYGDQGLN--LENPYIVNREMFVSKKRY-----MFYANVKYD--ILSLWNI 553
QY 474 -GRFKGDKLLPOK-----STIVOPAGS-----QYENTFYFDDAAL--KKDIYRLNY 516

Db 554 AGRIRDVNTWTSERKLHASTIKLHAQSDKGAYNRSMEEQYQOTYADIMLVNKNFGFNL 613
QY 517 STN-----TVGRFGGE-----YGYGSDDEPKRAFGENSPYTKKHCNQSCGI 560
Db 614 TANAGFSYEDHLTGTGMIGIGKLFVTPNLFSAI-----NFDPASGPGSQSHTHRNNSVFV 668
QY 561 -----YEPVLKYGKKRANNNHSVISADFGDYEMPFAFASYSKTRHMPNIQEMY----- 607
Db 669 STELGYSKMLYLTLTGROEWASQLVNSDQTYFYPSVGVSGV-----ISEMVSILPFI 723
QY 608 -----FSQIGDSGVHTALKP-----ERANTWQFGFNTYK 636
Db 724 WKMRASFAEVGGPINVTGLTGTVDPMKGGVINPISVYPPNFKAQOTKSYELGTN--- 780
QY 637 KGLLKQDDTLGLKLVGRYSIDN--YIHNVYKWDLNGNIPSWVSSTGLA-YTIQHRNF 693
Db 781 -----LRLFNKNINIDATVYLTDTYNQTF-----LSSMSPASGYSGFYVQ--- 820
QY 694 KDKVHKHGFELNYD--YGR--FFTNLASYOKS 724
Db 821 AGVRNKGIELSLGYNDRFGKVGATNLTITANRN 855
RESULT 9
US-09-540-209B-10058
; Sequence 10058, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540, 209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 10058
; LENGTH: 707
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-10058
Query Match 3.7%; Score 180.5; DB 5; Length 707;
Best Local Similarity 21.0%; Pred. No. 8.9e-06;
Matches 175; Conservative 101; Mismatches 332; Indels 225; Gaps 48;
QY 36 QIOVLEDVHVYKAKRVPKDKKVFVDARAVSTRQDIEKSSNDLVRSIPCAFTQODKSSG 95
Db 1 EVOI-----VSTRATSKTPVAFNTVSKUELKONF--GQDIPFLLSWTPSALTSDAGAG 53
QY 96 I--VSLNIRGDSGFRVNTWVDGITQT-----FYSTSTD-----AGRA-G 132
Db 54 IGYTTLVRGTDG-TRINITANGIPMNDASHTLFWNNMPDFASSVKDIOVQRCAGTSTN 112
QY 133 GSSQFGASVD-----SNFIAGLDVYKSGFSGSAGINSLAGSANLRTL 174
Db 113 GAGAFGASVNMQTEGISMOPYAEINASYDSFNAHKETVK--FGTGLLKDHAFDARLSTI 170
QY 175 GYDDVVQ--GNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESASGVLYGHSRRSVAQ 232
Db 171 GTDGYIDRASVDLYSYFAQGGYFADNTSVKFTFGKERTYHA-----WNATKEEMK 223
QY 233 NY--RVGGGGOHIGNFG-ABYLERRKQRYFVQEGGLKFNNS-SGKWERDFQRPYWK-TKW 287
Db 224 KYGPRENSCMYTDHGHITFYKQDQTDNYLQMNLYLLNHTFSAANNLNAALHYTKGDY 283
QY 288 YQYNDPQELQKIEGHDKSWRENLAPOYDITPIDPSSL-KOOSAGNFKLYDGVFN- 344
Db 284 YQYKEDRSLSKEY-----RLHPFMYDGKEVEKSDLIQKKMDNHFQ---GGVFSV 330
QY 345 KYTAQFEDLNTKGSRKIINRNQYFNGLSLNYSANLNTAAVNSGRQKYPKSKGTGWG 404
Db 331 NYPHQNDAS-----LVNA-----LNYDGHGFRGVIVWK--NYIG-E 365

Db 119 WGADH-GLELDA-FNAGQVSRKGFASLLYSGDAMGGAIELVPL---PLPAGNRLFGEAS 173
Qy 188 LLLGLTGTSTKGNAMAAIGARKWLESASGVLYGHSRESVAQYRVGGGGOHGNFG 247
Db 174 LLGKSVNGT--LGSLMLGKIKDAM-----YTWARYS-----EHOFGD 211
Qy 248 AE-----YLERKORYFVQEGGLKFNSSNGSKWERD-----FORPYWKTKWYQY 291
Db 212 IPTDTIVLTQMPVYHRR---LK---NTAGFERDVSWAAGFRKERYVSSVWVSNVQKT 265
Qy 292 N-----DPQELQYIEGHKSWRENLAPOYDITPIDPSSLUKQOSAGNLFKLEYDGV 342
Db 266 GFFPCAHCIPDVSRLQ-----DQDGRNIELPYSOVNHILKVSTRQSLLYDKWALTWDIG 319
Qy 343 FNK-YTAQFRDLNLIKGRKILINNYQFNGLSLNSVANLNLTAAYNSGRQKYPKGSKFT 401
Db 320 FOKNREWSRFHYDAQVPDQKDPKELAFNTLNTSSAVKLKLFASVWQHT-----A 374
Qy 402 GWLLKDFETYNNAKILDLNNTATFRLPRETELOTTLGFNYPHNEYGNRFPPEELGLFFD 461
Db 375 GWDV-----QYQRNTIAGYSFLLPAY--RRF--TTGAFW- 404
Qy 462 GPDQDNGLYSLGRFKGDKLLPKOKSTIVQPGAGQYNTFFDAAKKDIYRLNYSNTV 521
Db 405 -----MTYRPG-----PTLSFGGLRYDYCKIDASAYTD 434
Qy 522 GYRFGGEVGYGSDDEFKRAFGENSPTYKKHCHQSCG-----IYEPVLKYYKKRANHS 577
Db 435 PYL--AIYLRQGGDEFIRKYEWRSPVRRHFHFDYSGSLGWSP-----SGHL 483
Qy 578 VSISADFGDYMPPFASYSRTHRMPIQEMVFSQIGDSGVH-----TALKPERANT 627
Db 484 LQVNV-----GHSFRLPGANE-----LASGVHGHGTPRHFGQDAALASERG-- 524
Qy 628 WQGFEN-TYKGLLKQDDTLGLKLVGRSRIDNVIH-NVYGKWWDL--NGNIPSWVSGTG 683
Db 525 WQFDASYTYENG-----PLSVLSLSPVSWFSNYIFLRPTGEWILPHAGQIYRYTGA 578
Qy 684 LAYTIQHRNFKDKVHKHGFLELYDGRFF--TNSLYAYQKSTQ--PTNFSDAESP 737
Db 579 LF-----AGGEAAVGDILRHFNRYVSGEYVYTYNCDEHIIPLSFSPASLR 624
Qy 738 NNASKEDQLKOGYGLSRVSALPRDYGRLEVGTWRWLGNKLTGGAMRYFGKSIRATAERY 797
Db 625 NTLT-----WRYKEFSYGEVQVHIAAQHVRARNEDP 655
Qy 798 IDGT---NGGNTSVNRLGKRISIKQETTLARQPL-----IFDFYAA-----EPKKN--L 842
Db 656 TPGAQLLNAGVSANLRIGG---IWAETLSARNLSGAKYFNHLSFYRKYRVEIPEGRNFQI 712
Qy 843 IFRAEVKNL 851
Db 713 LIKVPFKSL 721

RESULT 12

US-09-540-209B-7143
; Sequence 7143, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Bretton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540, 209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 7143
; LENGTH: 833
; TYPE: PRT
; ORGANISM: B. fragilis
US-09-540-209B-7143

Query Match 3.6%; Score 177; DB 5; Length 833;
Best Local Similarity 19.4%; Pred. No. 2,1e-05;
Matches 184; Conservative 119; Mismatches 349; Indels 298; Gaps 42;

Qy 9 PICPYLMGVMLYHHYSYAEDAG-----RAGSEAQIQVLEDVHVRAKRV-----K 52
Db 51 PVAIAIEGV--YIGGYTNENGVIHNDVPTGSGTIVVSGIGVTKKVIPIHTAGKVNRP 108
Qy 53 DKVFTDARAVSTPDIFKSE-----NLDNIVRSIPGAFTOODKSSGIVSLNIR 102
Db 109 DIEIDTQAELEEVQVIGKSEARRQEQAYAISVDLTKKAYNSAAPLNKLLNNVSSVRIR 168
Qy 103 GDSGFG-----RVNTMVDGITTQTFYSTSTSDAGRAGSSGOFASVDSNFTAGL 149
Db 169 EGGMGSNFNSLNGFSGNQVKFELDGI PMDNF-----GSSFNLANSANMAERV 218
Qy 150 DVYKGSFSGAGINSLAGSANLRTGLGVDDVVQGNNTVGLLKLGTGNSTKGNMAAIGA 209
Db 219 EVYKGLVPLNLAGADALGGAVNIVRRDANLDATYSFG-----SFNTHKVSVNGA 268
Qy 210 RKWLESGASV--GVLYGHSRRSVAQNYRVGGGGOHGNFGAEYLERKORYFVQEGGLKF 267
Db 269 YTHLKTGFTVRANAFNYS-----DNDYKV-----FVPIIDLAT 302
Qy 268 NSNSGKWERDFORPYWKTKWYQKYNPDQELQKYIEG-----HDKSWRENLAPOYDIT 319
Db 303 NKKIDE-----RWVRRFND-----AYRSGGIRLETGITNPKPYADYLLAGIILS 345
Qy 320 PIDPSSLUKQOSAGNLFKLEYDGVFNKYTAOFRDLNTHIKGRKII-----NRNVOFNVGLS 374
Db 346 KND-----KDVQTGATMDAVYGGV-----KMKSESVIPSTRYKDKDLFLDGLS 388
Qy 375 LNSYANLNLTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDLNNTA--TERLPRETE 433
Db 389 LSLYG-----TNSVNTFNVDTIARRNWLGESV 417
Qy 434 LQTLTGFENYFHEYGKNRFPPEELGDFDQDQNGLYSYLGRFKGDKGLLPQKSTIVQPA 493
Db 418 PSISAGEGYTDSKIKNR--EWLG-----NGNISYV-----IDGHSLLNLHV 458
Qy 494 GSQYNTFYFDAAALKKDIYRLNYSNTVGYRFGGEYTYGYSDDDEFKR-----AFGE----- 545
Db 459 VSMARRTM-----NDKVRPDDENNPNVPOQLTKNITG--LGWQIRYDRWANVFGKMYKL 510
Qy 546 NSPTYKKHCHQSCGIYEPVLKKY-----CKKRANHSVSIADFGDYFMPFA-----SYS 595
Db 511 YSITYKR-----LDEYTENARWEKVRDHKTNFGYGAAATYIILPSLOAKESYE 558
Qy 596 RTHRMPIQEMVFSQIGDSGV---HTALKPERANTWQGFEN---TYKKGLLKQDDTLGLK 649
Db 559 HAYRLPESIEF---GDGLIQORNPDLKPSSRNLLNGLSFQTQGAHQLSADGNF--- 611
Qy 650 LVGYRSRIDNYIHNYYGKWWDLNAGNIPSVSVSTGLAYTIOHRNPKDKVKHKGFELELNYD 709
Db 612 --IYRYTTDFILKGV-----SLTSNPTTGYENLG-----KVLTKGVEAAVRYN 652
Qy 710 Y-GREFTNLSYAYOKSTOPTNFSDASES-----PNNASKEDQLKOGYGLSRVSALPRD 761
Db 653 YKOLFHTGAGTYQDITDRORYEKTDFSVEGEGITENITYKE-----RLPNIPYL 702
Qy 762 YGRLEVGT---WLGNKLTLLGAMRY-----GKSIRATAERYA--DGTNGG 804
Db 703 FANADAGVRPHDLJWRNSVLTFDYNLNYIHSYVLSFPLGAKSKKVIPIQFSDHALGY 762
Qy 805 NTSNVRLQKRSIKQETTLARQPLIFDFYAAYPEKKNLIFRAEVKNLFD 854
Db 763 SMDN-----KYSYVVECTNLTKOKLYDNYRLQKPGR--AFNVKLRYFFSK 806

RESULT 13

US-09-540-209B-8474
; Sequence 8474, Application US/09540209B

GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 8474
; LENGTH: 702
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-8474

Query Match 3.5%; Score 172; DB 5; Length 702;
Best Local Similarity 18.9%; Pred. No. 4.2e-05;
Matches 143; Conservative 107; Mismatches 267; Indels 240; Gaps 36;

Qy 170 NLRTGLVDDVVOGNNVTGGLLGLTGTNSTKGNMAAIGARKWLESASGVLYGHRRS 229
Db 95 SLNLYQSDAVK--HFAGVTVKDYGGIGGLKTVSLRSLGAEH-----TAVG--YDGITIS 145
Qy 230 VAQNYRVGGGQHIGNGAEYLER-----RKQRYVEQEGGLKFNNSGKWERD 277
Db 146 DCQT-----GQIDIGRSLDNRSLNSGSDNIFQPARFFASAGILNQTLT----- 194
Qy 278 FQRYWTKYQKYNPQELQYIEGHDKSWRENLAPOYDITP-----IDPSLKQOQSAGNL 334
Db 195 ---PQFK-----DNRRTNLSASFKTGSLVNPSPSLLEQLSRK 230
Qy 335 FKLEVDG-----VFNKYTAQFRLNTKIGSRKINRNQYNGL--SLNSYANLNL 383
Db 231 WLSANGEMWSADGHYFTLHYGDNLDLTREKRNKTEVKNLRAEAGLFGNFSDEQWRL 290
Qy 384 TAAVNSGRQKPKGSKTGWGLLKDFETYNNAKILDLN---NTATFRLPRETELOTLG 439
Db 291 KAYYQSSRGLPNATTY-----YDYSOHLWDKNVQSOYKKEFSQWQFQISAK 342
Qy 440 FNYPHNEYGNRFPPELGFDDPDQDNGLYSYLGRFKGDKGLLPQKSTIVQAGSQYFN 499
Db 343 WWSYQRY-----LD-PD-----YKSGEG-----KTENSYYQQ 369
Qy 500 TFYEDA-ALKKDIYRLAYSTNT-----VGYRFGGYTCYGGSDDEKR 541
Db 370 EYLSASALYRLVLSLFSUSTDASINRLNANLKDFAYPTRYSLWTAFAGKYVND----- 424
Qy 542 AFGENSPTYKHKHCQSGIYEPVLK---KYGKKRANHVS--ISADFGDYFMPFAS--- 593
Db 425 -----WLTASVLTATVINEEVROGSAANRRKLSPYVAS-----FKPFASEEP 469
Qy 594 -----YSRTHRMPNIQEMYSQIGDSGVHTALKPERANTWQFGNTYKGLLKQDDTLGL 648
Db 470 RIRLFYKIDFRLPSFNDLYYGQVG---NTNLKPESTTQYNLGL-TYSRSINELIPVSV 524
Qy 649 KLVGYRSRIDNYIINVGKWDNLGNTPSVSVSTGLAYTQHRNFKDKVHKHGFELNLY 708
Db 525 TADAYNKVKDKIIATPK-----NLFIW-SMVNLG-----KVDIKGDIAGNI 567
Qy 709 D---YGRFFTNLS--YAYQKSTQPTNFSASESPNNASKEDQLKQGYGLSRVSLPRDYG 763
Db 568 SLQWFKLRVNLNSNYIYQRAL-----DMTEPGKTYKQ-----QIATPRVSG 611
Qy 764 RLEVG--TRW--LGNKLTGGAMRYFGKSRATAEERYIDGTNGGNTSNVRQLGKRSTIKQ 819
Db 612 SGQAGIETPWNLVSYSLFSGKRYMLGNLR---ENRLDSYSDHSVSVSRDLRINVT 667
Qy 820 TETLAROPLIFDFYAAYPEKKNLIFRAEVKNLFDPRY 856
Db 668 SLTV-----EVLNLDKNY 681

RESULT 14

US-09-540-209B-9746
; Sequence 9746, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 9746
; LENGTH: 783
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-9746

Query Match 3.5%; Score 172; DB 5; Length 783;
Best Local Similarity 20.7%; Pred. No. 4.9e-05;
Matches 137; Conservative 84; Mismatches 258; Indels 184; Gaps 29;

Qy 34 EAQIQVLEDVHVAKRVPKDKKVFTRDARAVSTRQDIFKSSSENLDNIVRSIPCAFTQQDKS 93
Db 107 EPETQSLGEVVTAKSEARQLREQAMPMSVLSMQLOLQCTVSNVODVLSKTVGTVIRNTRGG 166
Qy 94 SGIVS-LNIRGDSGFRVNTMVDGITQTFYSTSTADAGAGSSQFGASVDSNFTAGLDV 152
Db 167 VGSRRSVVRGLEG-KRIGFFIDGSPMNDNSDFIDN-----DIPVDMIDRIEY 215
Qy 153 KGSFSGSAGINSLAGSANL-----RTLGVDDVVQGNNTYGLLLGLTGTNSTKGNAMA 205
Db 216 KGVTPARFGSGGVGAVNIVREYPPKYLDASYSIESFNTHKLSL-----VTKRNI-- 266
Qy 206 AIGARKWLESASGVLYGHSRRSVAQNYRVGGGQHTGNFGAEYLERKORYVQ---E 262
Db 267 --ATKGLFEGG--GGFTYS---DNNYKME-----SPEEGLIIKRNDHKKKLAVA 311
Qy 263 GGLFENSNKGW-----ERDFORPYWKWKYOKYND---POELQK---YI 301
Db 312 GSLK---ARKWFDLAEFPVFIHTFKEIOGIEYNIKAHTYSDAFTFANKLEKENFLT 367
Qy 302 EGHDKSWRENLAPOYDI--TPIDPSLSKQOQSAGNLFKL--EYDGVFNKYTAQFRLDNLTKIG 358
Db 368 EGLD--MESNLAYAYTVFHVWDTAARYNWDGTTYPAVSEYGEIGKWNARN----- 419
Qy 359 SRKIINRYQPNYGLSNLSYANLNLTAAYNSGRQKYPKGSKEFGW-----GL 405
Db 420 EKHTITHKLHLYVINNNHSINLSLFSFASGHPKDDLKNKVGVKTNFRSTMASIAGL 479
Qy 406 LKDEFTYNNAKILDNL-----NT-----ATFRLPRETELOTLGTFNYPHNE 446
Db 480 GYDFRTDNDIFLNSLNKYMYMGNTHTMSSTMSSEAEKVDMLKRDGFSNALRYRFTPDF 539
Qy 447 YGK-----NRFPEELGLFDDQDNGLYSLGRFKGDKGLLPQKSTIVQAGSQYFNT 500
Db 540 MGKLSUGVDVRLPAESELLGDG-----YTAPSGN--LLPERNTSV-----NL 580
Qy 501 FYFDAALKDITRLNYSNTVGYRFGGEYTYGSDDEKFRAGFENSPTYKHKHCQSGI 560
Db 581 GFLDLRTEKDNALQVEVNT-----FYGVLENMIRFTG-----GY 615
Qy 561 YEPVLKYYGKKRANHNSVSVISADFGDFMPFASV-----RTH-----RM 600
Db 616 LQSOYQNFGRMKTGLGVEVEVKADLTHWLYGYCNMTYQDLDRVKRFPENHTITNPTKGRM 675
Qy 601 PNI 603
Db 676 PNI 678

RESULT 15
US-09-584-501A-2
; Sequence 2, Application US/09584501A

THIS PAGE BLANK (USPTO)



INTERFERENCE

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:53:48 ; Search time 25.29 Seconds
(without alignments)
889.520 Million cell updates/sec

Title: US-09-762-926-6
Perfect score: 4894
Sequence: 1 MRSSFLKPCIFLGMVGLY.....SVLTNFAGRFLITMSYKF 921

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	276.5	5.6	915	3	US-08-613-009A-18
2	273.5	5.6	915	1	US-08-487-890A-96
3	273.5	5.6	915	2	US-08-363-124A-2
4	273.5	5.6	915	2	US-08-478-435-96
5	273.5	5.6	915	2	US-08-337-483-96
6	273.5	5.6	915	2	US-08-478-373-96
7	273.5	5.6	915	3	US-08-474-671-96
8	273.5	5.6	915	3	US-08-483-577A-96
9	273.5	5.6	915	4	US-08-897-438-96
10	273.5	5.6	915	4	US-08-637-654-96
11	269	5.5	908	1	US-08-487-890A-94
12	269	5.5	908	2	US-08-478-435-94
13	269	5.5	908	2	US-08-337-483-94
14	269	5.5	908	2	US-08-478-373-94
15	269	5.5	908	3	US-08-474-671-94
16	269	5.5	908	3	US-08-483-577A-94
17	269	5.5	908	3	US-08-488-194-4
18	269	5.5	908	3	US-08-613-009A-16
19	269	5.5	908	4	US-08-897-438-94
20	269	5.5	908	4	US-08-867-921-4
21	269	5.5	908	4	US-08-637-654-94
22	267.5	5.5	791	3	US-08-537-361E-4
23	264.5	5.4	791	4	US-08-817-707-4
24	261.5	5.3	790	4	US-08-817-707-6
25	256.5	5.2	911	1	US-08-487-890A-95
26	256.5	5.2	911	2	US-08-478-435-95
27	256.5	5.2	911	2	US-08-337-483-95

28	256.5	5.2	911	2	US-08-478-373-95	Sequence 95, Appl
29	256.5	5.2	911	3	US-08-474-671-95	Sequence 95, Appl
30	256.5	5.2	911	3	US-08-483-577A-95	Sequence 95, Appl
31	256.5	5.2	911	3	US-08-448-194-6	Sequence 6, Appl
32	256.5	5.2	911	3	US-08-613-009A-17	Sequence 17, Appl
33	256.5	5.2	911	4	US-08-897-438-95	Sequence 95, Appl
34	256.5	5.2	911	4	US-08-867-921-6	Sequence 6, Appl
35	256.5	5.2	911	4	US-08-637-654-95	Sequence 95, Appl
36	253.5	5.2	790	4	US-08-817-707-8	Sequence 8, Appl
37	253.5	5.2	909	2	US-08-363-124A-4	Sequence 2, Appl
38	252.5	5.2	884	1	US-08-066-167-2	Sequence 2, Appl
39	250.5	5.1	790	3	US-08-537-361E-8	Sequence 8, Appl
40	249	5.1	941	4	US-09-074-658-75	Sequence 75, Appl
41	247	5.0	792	3	US-08-537-361E-6	Sequence 6, Appl
42	246	5.0	792	1	US-08-326-670A-2	Sequence 2, Appl
43	246	5.0	792	3	US-08-990-470A-2	Sequence 2, Appl
44	246	5.0	792	4	US-08-817-707-2	Sequence 2, Appl
45	239.5	4.9	792	3	US-08-537-361E-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-613-009A-18
; Sequence 18, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,009A
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-542
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-613-009A-18

Query Match 5.6%; Score 276.5; DB 3; Length 915;
Best Local Similarity 21.2%; Pred. No. 4.2e-14;
Matches 221; Conservative 128; Mismatches 328; Indels 367; Gaps 57;

.Qy	5	FLKPCIFYLMGVTLYHYSVAEDADGAGSEAQIOVLVDHVVKAK---RVPDKKKVFTDARA	62
.Qy	6	FLRNLCLSLMTAL---PAYAENV---QAQ-EQEKQDITQVRAKKAKQKTRDNEVFTGLGKL	61
.Qy	63	VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK---SSGIIVSLNIRGDSGFGRVNTWVDGIT	118
.Qy	62	VKTADTLSKEQVLDIRDLTRYDPCIIVVEGRGASSG---YSIRG-MDKNRVSLTVDGLA	117
.Qy	119	Q-TFYSTSTDAQ---RAGGSQFASVDSNFIAGLDVVKVFGSGSAGINSLAGSANLRTLG	175
.Qy	118	QIOSYTAQAALGGTRTAGSGALNETEYENKVAEISKGSNVEQSGALAGSAVAFQKT	177
.Qy	176	VDDVQGNNTYIGLLLK-GLTGITNSTKGNMAAIGARKWLSCASGVLYGHSSRVSQAQY	234
.Qy	178	ADDVIGEGROWGTQSTAYSGKNRGLTQSLALAG-----	211
.Qy	235	RVGGG-----GOHIGNFGAEYLERRKORYFVQEGALKFN-----SDSGKWERDLOROQ	282
.Qy	212	RIGGAEALLIRTRHAGEITRAHEAAGR-----GVQSFNRLAPVDDGSKYAFIVEEE	266
.Qy	283	WKYKPKYKNYNOELQ---KYITEGHDK---SWRENLAPOYDITPIDPSSLKQOOS---A	330
.Qy	264	-----CKNGCHEKCANPRDVGVEDKROTVSTRDYTGPNRELA---DPLSVESRWSLFRP	316
.Qy	331	GNLF--KLEYDGVFNKYTAQ---FRDLNTKIGSRKII-----NRNYQFN--Y	370
.Qy	317	GFRFNKRHYIGGILERTOOTFTRDMTVPFAFLTKAVFDANQKQAGSLRGNGKYAGNHKY	376
.Qy	371	GLSLNPTNLNLTAAYNSGRQKYPKGSKTTGGLLKDFETYNNAKILDLNNTATFRLPRE	430
.Qy	377	G-----GLTSGENNAVPAQAEY-GTGVFYD-----	401
.Qy	431	TELQTLTLGFNYFHNEXKGNRFPBELGLFDP---GPOODNGL-----	468
.Qy	402	THKRSYGLVEYVYNADKDTWADYARLSYDROGILGDNHFFQOCHSADGSKYCRPSADK	461
.Qy	469	-YSYLRFPKDGKLLPKOKSTIVOPAGSOYNTFYFDAALKKDIYRLNVTSTNTVGY-RFGG	526
.Qy	462	PFSY---YKSDRYIYGESHKLLQAAPKPKSFDATAKI-----RHNLSVN-LGYDRFGS	508
.Qy	527	-----EYTGYYGSDDEFKRAFGNSP-----	547
.Qy	509	NLRHODY--YYQSAN---RAYSKTTPQNNNGKKTSPNGREKNPYWVSTGRNVVTRQICL	563
.Qy	548	-----TYKKHCNRSC---GIYEPVLKK-----YGKRRANNHSVS-----	578
.Qy	564	FGANTYTDCTPRNGSKSYAAVRONRVRLGRWADVGAGLRYDRSTHSDGDSVSTGTHRT	623
.Qy	579	ISADFGDYFPFASYSRTH-----RMPNIOEYFISOICDGSVHTALPERANTWQFG--	630
.Qy	624	LSWNAGVLVLPADWLDTYVSTGFRPLPSFAEYGMWSGDKIAKVAKIDPEKSFNKEAGIV	683
.Qy	631	-----ENTYKKGILLKQDDPLGLKLVGYRSRIDN-----YTH-----	661
.Qy	684	FKGDFGNLEASWENNAYROLIVR-----GYEAQDKGKEQVKGPNPAYLNAQSARIT	734
.Qy	662	--NVIYK--W---WDLNGDIP-SWVSSTGLATV-TOHRNFKDKVHKHKGFELELNYDYGRF	712
.Qy	735	GINILKIDWNGVWD---KLPEGWYST--FAYNRVVRDIKKRADRTDITQSHL-FD----	784
.Qy	713	FTNLVAYQKSTOPTNFDASSESNNASKEDQLKQYGLSRVSALPRDYGRLEVGTRWLG	772
.Qy	785	-----AIQPSRYV-----VGSYDQPEKGWNGMLT-----YSKAKETELLG	823
.Qy	773	NKLTGGMAMRYFGKSRATAEERYIDGTNGGNTSNFRQLGKRKSIKOTETLAROPILIFOFY	832
.Qy	824	SRALLNG-----NSNRTATSRRTFPWIVDVS	851
.Qy	833	AAYEPKNLIFRAEVKNLFDRIYI	856
.Qy	852	GYTTVKKHETLRAGVYNLLNHRVY	875

```

RESULT      2
US-08-487-890A-96
; Sequence 96, Application US/08487890A
; Patent No. 5708149
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,890A
; FILING DATE: 07-JUN-1993
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-466 MS:j:b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-487-890A-96

```

Query Match 5.6%; Score 273.5; DB 1; Length 915;
Best Local Similarity 21.3%; Pred. No. 7.5e-14;
Matches 22; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

Qy	5	FRUKPCFYLMGVTLHYHYSAEDAGRAGSEAQTQVLEDVHVAK---RVPKDKKVFYTDARA	62
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	7	FRNILLCSLMTAL---PAYAENV-QAG-QAQPQKJDTIQVAKKQKTRDNEVTGLGKL	61
Qy	63	VSTRQDIFKSSE-NLDNIIVRSIPG-AFTQODK---SSGIIVSLNIRGDSGFRVNTVMVDGIT	118
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	62	VKTADTLSEQVLIIDRLTYDPIAVFQGRGASSG---YSIRG-MDKNRVSLTVVDGLA	117
Qy	119	Q-TFYSTSTDAG---RAGGSSQFGASVDSNFIAGLDVYVQFSFGSAGINSLAGSANLRTLG	175
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	118	QIQSYTAQAALGGTRTAGSSGAINEIYEYENKRAVEISKGSNVEQSGSALAGSVAFOYKT	177
Qy	176	VDDVQGNNTYGLLLKGLTGNTSTKGNMAAIGARKWLKESGASGVGLYIGHSRRSVAQNTRY	235
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	178	ADVIVGEGROWGT-----QSKTAYSGKNRGLI-----TQSIALAGR	212

```
QY 236 VGGG-----GQHNFGAEYLERRKORYFVQEGALKFN-----SDSGKWERDLQROOW 283
Db 213 IGAEALLIIRTRHAGEIRAEHAGR-----GVSFNRFLAPVDDGSGYAYFIVEEE- 263
QY 284 KYKPYKNYNNOEL-----QKYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AG 331
Db 264 ----CKNGGHECKKANPKDVGEDKQOTVSTRDYTGPNRFLA--DPLSYESRSWLFRPG 317
QY 332 NLF--KLEYDGVFNKTAQ---FRDLNFKIGSRKII-----NRRYQFN--YG 371
Db 318 FRENKRHYIGGILERTQOTFDTRDMTPVAPFLTKAVFDANQKAGSLRGNGKYAGNHKG 377
QY 372 LSLNPTNLNLTAAYNSGRQKYPKSGKFTGWLKDKDFETYNNAKILDNLNNTATFRPRET 431
Db 378 -----GLFTSGGENNAPVGAAY-ETGVFYD-----ET 402
QY 432 ELQTTLGFNFHNEYGNRPPEELGLFFD--GPDQDNGL----- 468
Db 403 HTKSRYGLEVVYTNADKDTWADYARLSYDRQGIGLGNHFFQTHCSADGSDKYCRPSADKP 462
QY 469 YSYLGRFKGDKLLPOKSTTVQAGSQYFNTFFYDAAKKDIYRLNYSNTVGY-RFGG- 526
Db 463 FSY---YKSDRVYIGESHKLLQAAFKKSFDTAKI-----RHNL SVN-LGYDRFGSN 509
QY 527 ---EXTGYGSDDEKRAFGENSP----- 547
Db 510 LRQDY--YQSAN---RAYSLKTPPNNGKKTSPNGREKNPYWSTIGRGNVVTQICLF 564
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YGKKRANNHSVS-----I 579
Db 565 GNNTYDCTPRSNKRSYAAVRDYNVRLGRWADVAGLRYDYRSTHSDGSGVSTGHTL 624
QY 580 SADFQGYFMPFASYRTH-----RMPNIQMYFSQIGDSGVHTALKPERANTWQFG--- 630
Db 625 SWNAGIVLKPADWLDLTYRTSTGRFLPSFAEMYGWRSQDKKAVKIDPEKSFNKEAGIVF 684
QY 631 -----FNTYKGLLKQDDTLGLKLVGYRSRDN-----YIH----- 661
Db 685 KGFNLEASFNNAVRLDLVR-----GYEAQIKGQKEQVGNPNAQASARITG 735
QY 662 -NVYK--W---WDLNGDIP-SWVSSTGLAYT-IQRNFKDKVHKHGFLELNDYDGRFF 713
Db 736 INILGKIDWGVND---KLEPGWYST--FAYNRVVRDINKRADRTDIQSHL-FD----- 784
QY 714 TNLAYAKSTQTPNFSADASESPNNASKEDQLKGYLSRVSALPRDYGRLEVGTWRLGN 773
Db 785 -----AIQPSRYV-----VSGYDQPEGKGVNGMLT-----YSKAKETELLGS 824
QY 774 KLTIGGAMRYFGKSIRATAERYIDGTNGGNTSNFRLGKRSTKQETLARQPLIFDYA 833
Db 825 RALLG-----NSRNTKATARRTPWYIVDVSG 852
QY 834 AYEPKKNLIFRAEVKNLFRDRIY 856
Db 853 YTVVKKHFTLRAGVYNLLNRYV 875

RESULT 3
US-08-363-124A-2
; Sequence 2, Application US/08363124A
; Patent No. 5912336
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
```

```
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,124A
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,254
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-363-124A-2

Query Match 5.6%; Score 273.5; DB 2; Length 915;
Best Local Similarity 21.3%; Pred. No. 7.5e-14;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRLLPKPICFLYMLGVTLYHYSYAEDAGRAGSEAAQIOVLDEVDHVKAK--RVPKDKKVFYTDARA 62
Db 7 FRNLILSLMTAL---PAYAENV-QAG-QAQEKQLDTIOVKAKKOKTRDRNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRVNTWVDGIT 118
Db 62 VKTADTLKSKQVLDIRDLTRYDFGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TFYSTSTDAQ--RAGGSSQFGASVDSNFTAGLDVYKGSFGSGAGINSLAGSANLRTL 175
Db 118 QIOSYTAQAALGGTRTAGSSGAINIEYENKAVEISKGSNSVEQSGGALAGSAVQTKT 177
QY 176 VDDVVOGNNTYGLLLKGLTGNTSTKGNMAAIGARKWLESGASVGVLYGHSSRVAQNYR 235
Db 178 ADVIGBGRGWI-----QSKTAYSGKNRGL-----TQSIALAGR 212
QY 236 VGGG-----GQHIGNFGAEYLERRKORYFVQEGALKFN-----SDSGKWERDLQROOW 283
Db 213 IGAEALLIIRTRHAGEIRAEHAGR-----GVSFNRFLAPVDDGSGYAYFIVEEE- 263
QY 284 KYKPYKNYNNOEL-----QKYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AG 331
Db 264 ----CKNGGHECKKANPKDVGEDKQOTVSTRDYTGPNRFLA--DPLSYESRSWLFRPG 317
QY 332 NLF--KLEYDGVFNKTAQ---FRDLNFKIGSRKII-----NRRYQFN--YG 371
Db 318 FRENKRHYIGGILERTQOTFDTRDMTPVAPFLTKAVFDANQKAGSLRGNGKYAGNHKG 377
QY 372 LSLNPTNLNLTAAYNSGRQKYPKSGKFTGWLKDKDFETYNNAKILDNLNNTATFRPRET 431
Db 378 -----GLFTSGGENNAPVGAAY-ETGVFYD-----ET 402
QY 432 ELQTTLGFNFHNEYGNRPPEELGLFFD--GPDQDNGL----- 468
```

Db 403 HTKSRYLEVYVTTWADKTDWADYARLSYDRQIGLDNHFQOQTHCSADGSKYCRPSADKP 462
QY 469 YSILGRFKGDKLPPQKSTVQVAGSYFNTFYFDDAALKKDIYRLNVTNTVGY-RFGG- 526
Db 463 FSY---YKSDRVYIGESHKLLQAAFKSFTAKI-----RHNL SVN-LGYDRFGSN 509
QY 527 ----EYTGYYGSDDEFKRAFGNSP----- 547
Db 510 LRHODY--YYQSAN---RAYSLKTPPNQNGKKTSPNGREKNPYVWSTGRGNVTVTRQICLF 564
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YKKRANNHSVS-----I 579
Db 565 GNNYITDCTPRSLNGKSYAAVRDNVRLGRWADVAGLRYDYRSTHSDGCVSTGTHRTL 624
QY 580 SADGDFYMFASFYSRTH-----RMPNIQMYFSQIGDSGVHTALKPERANTWOFQ--- 630
Db 625 SWNAGIVLKPADWLDLTYRTSTGRFLPSFAEMYGWSRSDKIKAVKIDPEKSFNKEAGIVF 684
QY 631 -----FNTYKGLKLQDDTLGLKLVGYRSRDN-----YIH----- 661
Db 685 KGDFGNLEASWFNAYRDLIVR-----GYEAQIKDGKEQVKNPYNQAQSARITG 735
QY 662 -NYGK-W---WDLNGDIP-SWVSSTGLAYT-IOHNFKDKVHKHGFELANVDYGRFF 713
Db 736 INILGKIDWGVWD---KLPEGWYST--FAYNRVRVRIKRRADRTDIQSHL-FD----- 784
QY 714 TNLAYQKSTQPTNFSDASESPNNAKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLG 773
Db 785 -----AIQPSRYV-----VGSYGQPEKGWGVNGLT-----YSRAKEITELG 824
QY 774 KLTGGAMRYFGKSIRATAERYIDGTNGGNTSNFRLGKRKSQKTTTLARQPLIFDFA 833
Db 825 RALLNG-----NSRNTKATARRTPWYIVDVSG 852
QY 834 AYEPPKNLIFRAEVKNLFDRIY 856
Db 853 YYIVKKHFTLRAGVYNLLNRYV 875

RESULT 4
US-08-478-435-96
; Sequence 96, Application US/08478435
; Patent No. 5922323
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,435
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-478-435-96

Query Match 5.6%; Score 273.5; DB 2; Length 915;
Best Local Similarity 21.3%; Pred. No. 7.5e-14;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRLKPCIFGLMGTLYVHYSAEDAGRAGSEAOIQVLEDVHVAK--RVPKDKKVFETDARA 62
Db 7 FRNLICLSLMTAL---PAYAENV-QAG-OAQEKQLDTIOVKAKKQKTRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGVRVMTVDGIT 118
Db 62 VKTADTLSRQVLDIRDLTRYDPCIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDG 117
QY 119 Q-TFYSTSTTAG--RAGSSQFGASVDSNFIAGLDVVVKGSFSGSAGINSLAGSNLRTLG 175
Db 118 QIOSYTAQAALGCTRTAGSSGAINIEYENVKAVEISKSGNSVEQSGGALAGSVAFOTKT 177
QY 176 VDDVQGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGLVYHSRVSVAQNYR 235
Db 178 ADDVIGEGROWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212
QY 236 VGGG-----GQHIGNFGAEYLERRRQRYVQEGALKFN-----SDSGKWERDLQKQW 283
Db 213 ICGAEALLIRTRHAGEIRAEHAAGR-----GVOSFNRLAPVDDGSKYAFIVEEE- 263
QY 284 KYPKYKNYNQEL---QKYLEGHDK---SWRENLAPOVDITPIDPSSLKQOS-----AG 331
Db 264 ----CKNGGHEKCKANPKDVGDGDKRQTVSTRDYTGPNRFLA--DPLSYESRSLWFRPG 317
QY 332 NLF--KLEVDGVFNKYTAQ---FRDLNLTIGSRKII-----NRRYQFN--YG 371
Db 318 FRENKRHYIGGILERTQQTFTDFTRDMTVPAFLTKAVFDANQKQAGSLRGNGKTAGNHKY 377
QY 372 LSLNPTYNLNLTAAYNSGRQYKPKSKFTGWLGLLKDFETYNNAKILDLNNTATFRLPRET 431
Db 378 -----GLFTSGENNAFVGAAY-CTGVFYD-----ET 402
QY 432 ELQTTLGFNFHNEYKKNRPEELGLFFD--GPDQONGL----- 458
Db 403 HTKSRYLEVYVTTWADKTDWADYARLSYDRQIGLDNHFQOQTHCSADGSKYCRPSADKP 462
QY 469 YSILGRFKGDKLPPQKSTVQVAGSYFNTFYFDDAALKKDIYRLNVTNTVGY-RFGG- 526
Db 463 FSY---YKSDRVYIGESHKLLQAAFKSFTAKI-----RHNL SVN-LGYDRFGSN 509
QY 527 ----EYTGYYGSDDEFKRAFGNSP----- 547
Db 510 LRHODY--YYQSAN---RAYSLKTPPNQNGKKTSPNGREKNPYVWSTGRGNVTVTRQICLF 564
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YKKRANNHSVS-----I 579

Db 565 GNNTYDCTPRSRNGKSYAAVRDNVRLGRWADVAGAGLRVDYRSTHSDGVSSTGTHRTL 624
QY 580 SADFGDYFMPFASYSRTH-----RMPNIQEMFYFSQIGSGVHTALKPERANTWQFG--- 630
Db 625 SWNAGIVLKPADWLDLTYSRTSTGFLPSPFAEMYGWSGDKIKAVKIDPEKSPNKEAGIVF 684
QY 631 -----FNTYKKGLLKQDDTLGLKLVGYRSRDN-----YIH----- 661
Db 685 KGDFGNLEASWFNNAYRDLIVR-----GYEAQIKDGKEQVKNPAYLNAQSARITG 735
QY 662 -NVYKG--W---WDLNGDIP-SWVSSSTGLAYT-IQHRNFKDKVHKHGFLELNDYGRFF 713
Db 736 INILGKIDWNGVWD---KLPEGWYST--FAYNRVRVDIKKRADRTDIQSHL-FD----- 784
QY 714 TNLISAYQKSTQPTNFSADASESPNNASKEDQLKGQGLSRVSPALPRDYGRLEVGTWRLGN 773
Db 785 -----AIQPSRYV-----VSGYDQPEGKGVNGMLT-----YSKAKEITELLS 824
QY 774 KLTGGAMRYFGKSIRATAEERYIDGTNGNTSNFRLGKRSIKQTETLARQPLIFDYA 833
Db 825 RALLNG-----NSRNTKATARRTPWYIVDVS 852
QY 834 AYEPPKNLIFRAEVKNLFDRIYI 856
Db 853 YTVVKKHETLRAGVYNLLNHRIV 875
RESULT 5
US-08-337-483-96
; Sequence 96, Application US/08337483
; Patent No. 5923562
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-410 MIS-jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-337-483-96

Query Match 5.6%; Score 273.5; DB 2; Length 915;
Best Local Similarity 21.3%; Pred. No. 7.5e-14;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;
QY 5 FRLKPCFYLMGYLYHYSYAEDAGRAGSEAQIOVLDEHVHAK--RVPDKKYFTDARA 62
Db 7 FRNLILCLSLMTAL---PAYAENV-QAG-QAQRKQLDITIQKAKKQKTRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRVNTMVDGIT 118
Db 62 VKTADTLKSKQVLDIRDLTRYDPGIAVVEQGRGASS---YSIRG-MDKNRVSLTVDLGA 117
QY 119 Q-TFYSTSTDAQ--RAGGSQFQASVDSNFIAGLDVVVKGFSFGSAGINSLAGSANLTGL 175
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKSGNSVEQSGGALAGSAVAFQTKT 177
QY 176 VDDVQGNNTYIGLLKGLTCTNSTKGNMAAIGARKWLESGASGVLYGHSRRSVAQNYR 235
Db 178 ADDVIGEGROWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212
QY 236 VGGG-----GOHIGNFGAEYLERRKORYFVOEGALKFN-----SDSGKWERDLQROW 283
Db 213 ICGAEALLIRTGRHAGEIRAHEAAGR-----GVQSFNRLAPVDDGSKYAVFIVEEE- 263
QY 284 KYKPYKNYNNQEL---OKYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS----AG 331
Db 264 ---CKNGGHEKCANPKDQVGDGKRQTVSTRDYTGPNRFLA--DPLSYESRSLFRPG 317
QY 332 NLF--KLEYGDFVFNKYTAQ---FRDLNLTIGSRKII-----NRNQQFN--YG 371
Db 318 FRFNKRHYIGGILERTQOTFDTRDMTVPAFLTKAVFDANOKQAGSLRGNGKAGNKKYG 377
QY 372 LSLNPTNLNLTAAYNSGROKYPKSGKFTGGLIKDPETYNNAKILDLNNTATFRLPRET 431
Db 378 -----GLFTSGENNAVPVGAIEY-GTGVFYD-----ET 402
QY 432 ELQTTLGFNFHNEYGKNRFPBELGLFFD--GPDQDNGL----- 468
Db 403 HTKSRYGLEYVYTNADKDTWADYARLSYDRQGIGLGNHFOOHCSDGSDKYCRPSADKP 462
QY 469 YSYLGRFKGKGLLPKOSTIVQAGSOYFNTFYFDAALKKDIYRLNYSNTVGY--RFGG- 526
Db 463 FSY---YKSDRVYIGESHKLLQAAFKKSFTAKI-----RHNLSVN-LGYDRGSGN 509
QY 527 -----EYTGYYGSDDEFKRAFGENSP----- 547
Db 510 LRHQDY--YYQSAN---RAYSLKTPPNNGKKTSPNGREKNPYWVSGRGNVTVRQICLF 564
QY 548 ---TYKKHCNRSK---GIYEPVLKK-----YCKKRANNHSVS-----I 579
Db 565 GNNTYDCTPRSRNGKSYAAVRDNVRLGRWADVAGAGLRVDYRSTHSDGVSSTGTHRTL 624
QY 580 SADFGDYFMPFASYSRTH-----RMPNIQEMFYFSQIGSGVHTALKPERANTWQFG--- 630
Db 625 SWNAGIVLKPADWLDLTYSRTSTGFLPSPFAEMYGWSGDKIKAVKIDPEKSPNKEAGIVF 684
QY 631 -----FNTYKKGLLKQDDTLGLKLVGYRSRDN-----YIH----- 661
Db 685 KGDFGNLEASWFNNAYRDLIVR-----GYEAQIKDGKEQVKNPAYLNAQSARITG 735
QY 662 -NVYKG--W---WDLNGDIP-SWVSSSTGLAYT-IQHRNFKDKVHKHGFLELNDYGRFF 713
Db 736 INILGKIDWNGVWD---KLPEGWYST--FAYNRVRVDIKKRADRTDIQSHL-FD----- 784
QY 714 TNLISAYQKSTQPTNFSADASESPNNASKEDQLKGQGLSRVSPALPRDYGRLEVGTWRLGN 773
Db 785 -----AIQPSRYV-----VSGYDQPEGKGVNGMLT-----YSKAKEITELLS 824
QY 774 KLTGGAMRYFGKSIRATAEERYIDGTNGNTSNFRLGKRSIKQTETLARQPLIFDYA 833
Db 825 RALLNG-----NSRNTKATARRTPWYIVDVS 852

STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US/08/483,577A
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-511
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-577A-96

Query Match 5.6%; Score 273.5; DB 3; Length 915;
Best Local Similarity 21.3%; Pred. No. 7.5e-14;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRLPICFVLMGVLYHYSAEDAGRAGSEAQIOVLEVDVHVAK--RVPKDKVFTDARA 62
DB 7 FRNLICSLMTAL---PAYAENY-QAG-QAQEKQLDTIOVKAKKQKTRRDNEVTGLGKL 61
QY 63 VSTRDIFKSSSE-NLDNLVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGVRNTMVDGIT 118
DB 62 VKTADTISKEQVLDIRLDTRYDPIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TFYSTSTADAG--RAGSSQFGASVDSNFTAGLDVYKGFSGSAGINSLAGSANLRTL 175
DB 118 QIQSYTAQAALGGTRTAGSGAINIEIEYENVKAVEISKSGNSVEQSGGALAGSVAFTKT 177
QY 176 VDDVVQGNNTYGLLLKGLTGNTSTNGMAAIGARKWLESGASVGVLYGHSRRSVAQNYR 235
DB 178 ADDVIGEGROWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212
QY 236 VGGG-----GQHIGNFGAEYLERRRKQRFVQEGALKFN-----SDSGKWERDLQROQW 283
DB 213 IGGAEALLIRGRAGEIRAHEAAGR-----GVOSFNRLAPVDGSKYAYFIVEEE- 263
QY 284 KYKPYKNVNNQEL-----OKYTEGHDK-----SWRENLAPOVDITPIDPSSLKQOS-----AG 331
DB 264 ---CKNGGHEKCKANPKDVGEDKQTVSTRDYTGPNRFLA--DPLSVESRSWLFPRG 317
QY 332 NLF--KLEYDGVFNKYTAQ---FRDLNKTGSRKII-----NRNQFN--YG 371
DB 318 FRFENKRYHIGGILERTQQTFTDMDTVPAPLTKAVFDANOKQAGSLRGKGYAGNHKYG 377
QY 372 LSLNPNYTNLNTAAYNSGRQKYPKSGKFTGWGLLKDFETYNNAKILDLNNTATFRLPRET 431
DB 378 -----GLFTSGENNAVGAEEY-GTGVFYD-----ET 402

QY 432 ELQTTLGFNFYHNEYGKNRFPPEELGLFFD--GPDQDNGL----- 468
DB 403 HTKSRYGLEIYYTNNADKDTWADYARLSYDRQIGLDNHFQOOTHCSADGSKYCRPSADKP 462
QY 469 YSILGRFKGDKGLLPQKSTIVQAGSQYFNTFYFDAALKKDIYRLNLTSTNTVGY-RFGG- 526
DB 463 FSY---YKSDRVYIGESHKLLQAAAFKKSFTAKI-----RHNL SVN-LGYDRFGSN 509
QY 527 ----EYTGYYGSDDEFKRAFGENSP----- 547
DB 510 LRHODY--YYOSAN---RAYSLKTPPONNGKKTSPNGREKNPYWVSGRGVNVTRQICLF 564
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YGKKRANNSVS-----I 579
DB 565 GNNYTTDCTPRISINGKSYAAVRDNVRLGRWADVAGLRYDYRSTHDDGGSVSTGTHRTL 624
QY 580 SADFGDYEMFPFASYSRTH-----RMPNIQEMYSQIGDSGVHTALPERANTWQFG--- 630
DB 625 SWNAGIVLKPADWLDLTYRTSTGFLPSPAEMYGWRSGDKIKAVKIDPEKSFNKEAGIVF 684
QY 631 -----FNTYKKGILLKQDDTLGLKLVGYRSRDN-----YIH----- 661
DB 685 KGDFGNLEASWENNAYRDLIVR-----GYEAQIKDKGEQVKGNPAYLNAOSARITG 735
QY 662 -NVYKG--W---WDLNGDIP--SWVSSSTGLAYT-IQHNRNFKDKVHKHGFELNDYDGRFF 713
DB 736 INILGKIDWNGVMD---KLPEGWYST--FAYNRVVRDIKKRADRTDIQSHL-FD----- 784
QY 714 TNLISAYOKSTQPTNPSDASESPNNASKEDOLKOGYGLSRVSAALPRDYGRLEVETRWLGN 773
DB 785 -----AIQPSRYV-----VGSQYDQPEKGWNGMLT-----YSKAKEITELLS 824
QY 774 KLTILGAMRYFGKSIRATAEERYIDGTNGNTSNFRQLGKRSIKQETLARQPLIFDYA 833
DB 825 RALLNG-----NSRNTKATARTRPWTIVDVSG 852
QY 834 AYEPPKKNLIFRAEYKKNLFDRIYI 856
DB 853 YVTVKKHFTLRAGVYNLLNHRV 875
RESULT 9
US-08-897-438-96
; Sequence 96, Application US/08897438
; Patent No. 6262016
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,438
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

: APPLICATION NUMBER: US 08/483,577
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/337,483
: FILING DATE: 08-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/175,116
: FILING DATE: 29-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/148,968
: FILING DATE: 08-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-720
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 96:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 915 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-897-438-96

```

	Query Match	5.6%	Score 273.5;	DB 4;	Length 915;
	Best Local Similarity	21.3%;	Pred. No. 7.5e-14;		
	Matches 222;	Conservative 129;	Mismatches 327;	Indels 365;	Gaps
Qy	5	FLRKPICFYLMGVTLYHYSAEADAGRAGSQAOTQVLEDYHVHKA--RVPKDKXVFVDARA	62		
Db	7	FRUNILCSLMTAL---PAYAENV-QAG-QAQEKQLDTQVKAKKOKTRRDNEVTGLGKL	61		
Qy	63	VSTRQDIFKSE-NLDNIVRSIPG-AFTQDQK--SSGIIVSLNRIGDSGFGRVNTWVDGIT	118		
Db	62	VKTADTLSKEQVLDRDLTRYDPGIAVVEOGRASSG---YSIRG-MDKNRVSLTVDDGLA	117		
Qy	119	Q-TFYSTSTDAQ--RAGSGSQFGASVDSNFIAGLDVVKVGSFGSAGINSLAGSANLRTLG	175		
Db	118	QIQSYTAQAALGGTGTAGSSGAINEIYEYNKVAAIEYISKGNSVEQSGGALAGSVAFQT	177		
Qy	176	VDDVVOCNNITYGLLLGLTCTNSTKGNMAAIICARKWLSEGSASGVLVYGHHSRRSVQA	235		
Db	178	ADDVIGEGROWGL-----QSKTAYSGKNRGL-----TQSIALAGR	212		
Qy	236	VGGG-----GOHIGNFGAEYLERRKORYFVQEGALKFN-----SDSGKWERDLQRQOW	283		
Db	213	IGGAELLIRTGRHAGEIRAHAAGR-----GVQSFNRLAPVDDGSKYAFIVEEE-	263		
Qy	284	KYPKYNNNOEL-----OKYIEGHDK-----SWRENLAPOVDITPIDPSSLUKQOS----	331		
Db	264	-----CKNGHEKCANPKKDVGDEKRQTVSTRDYTPGNRFLA--DPLSYESRSWLFPRG	317		
Qy	332	NLF--KLEYDGVENKYYTAQ--FRDLNTKIGSRKII-----NRRNQFN--YG	371		
Db	318	FRENKRHYIGGILERFOQTFRDMTVPAFLTKAVFDANQKQAGSLRGNGKTAGNHXYG	377		
Qy	372	LSLNPVNLNLTAAYSNGROKYPEKSGKFTGWGLLKDFETYNNNAKILDNLNNTATFLPRET	431		
Db	378	-----GLFTSGENNAVPVGAEY-GTGIFYD-----ET	402		
Qy	432	ELOTTIGFNFMNEYKKNRPPEELGFFD--GPDQONGL-----	468		
Db	403	HTKSRYGLEVVYTNADKDTWDARYLSYDRQGTYGLDNHQFOOTHCSADGSKDYCRPSADKP	462		
Qy	469	YSYLGRFKBGLLPKSTIVQPAGSOYFNFTFYFDAALKDIYRLANYSTNVGY-RFGG-	526		
Db	463	FSY---YKSDRVYIGSHKLUQAFAKSPDTAKI-----RHNL SVN-LGYDRFGSN	509		
Qy	527	----EYTGYGSDDEFKRAFGENSP-----	547		

```

510 LRHODY--YYOSAN---RAYSLKTPPNQNGKKTSPNGREKNPWWVSIGRGNVVVTRQICLF 566
QY ---TYKKHCRNSC---GIYEPVLKK-----YCKKRANHNSVS-----I 579
565 GNNIYTDCTPRSINGKSYAAVRDNVPLGRWADVGAGLRDYDRSTHSDGGSVSTGTHRTL 624
580 SADRGDYFMPFASYSRPH-----RMPNIOEMTFSGIGDSGVHTALKPERANTWQFG--- 630
625 SWNAGIVLKPADWLIDLTYRTSTGFLRPSFAEMYCWRSGDKIKAVKIDPEKSFNKEAGIVF 684
631 -----FNYYKGLLKQDDTLGLKLVGYRSRDN-----YIH----- 661
685 KGDFGNLEASWFNNAYRDLIVR-----GYEAQIKGQKEQVKNPAYLNAQSAKITG 735
662 -NVYGK--W---WDLNGDIP-SWVSSTGLAYT-IQHRNFMDKVHKHGFLELNYDYGRFF 713
736 INILGKIDWNGVD---KLPEGWYST--PAYNVRVRDIIKKRADRTDIQSHL--FD----- 784
714 TNLSYAYOKSTOPTNFSDASESPNNASKEDQLQKGGLSRVSAIPRDRDYGRLEVTGRWLGN 773
785 -----AIQPSRYV-----VGSYDQPEKGWNGMLT-----YSKAKEITELLS 824
774 KLTGGAMRYFGKSIRATAEBERYIDGTNGGNTSFNRQLGKRISIKOTETLARQPLIFDEYA 833
825 RALLNG-----NSRNTKATARTRPWYIVDVSG 852
QY 834 AYEPEKKNLIFRAEVKNLFRRYI 856
DQ 853 YTYVKKHFTLRAGVYNLLNHRVY 875

RESULT 10
US-08-637-654-96
; Sequence 96, Application US/08637654
; Patent No. 6358727
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Harkness, Robin E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew D
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,654
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA94/00616
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-595
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 96:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-637-654-96

Query Match 5.6%; Score 273.5; DB 4; Length 915;
Best Local Similarity 21.3%; Pred. No. 7.5e-14;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRLKPICFYLMGVLYHYSAEDAGRAGSEAIQVLEDVHVAK--RVPKDKKVFDTDARA 62
DB 7 FRLNILCLSLMTAL---PAYAENV-QAG-QAQEKQLDTIQVAKAKKQTRRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-ARTQDK--SSGIIVSLNIRGDSGFRVNTWVDCIT 118
DB 62 VKTADTLSEKQVLDIRLTRYDPIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TFYSTSTDAQ--RAGSSOFGASVDSNFTAGLDVYVKGFSGAGINSLAGSANLRTL 175
DB 118 QIQSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKGSNSVEQSGALAGSVAFOKT 177
QY 176 VDDVVQGNNTYGLLLKGLTGNTSTKGNAMAIGARKWLESGASGVLYLGHSSRRSVAQNYR 235
DB 178 ADDVIGEGROWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212
QY 236 VGGG-----GQHIGNFGAEYLERRKQYFVQEGALKEN-----SDSGKWERDLQOQW 283
DB 213 IGGAEALLIRGRHAGEIRAEAGR-----GVQSFNRLAPVDGSGYAYFIVEEE- 263
QY 284 KYKPKYNNQOEL-----QKYEIGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AG 331
DB 264 ---CKNGHEKCKANPKDVGEDKQTVSTRDYTGPNRFLA--DPLSYESRSWLFPPG 317
QY 332 NLF--KLEYDGVFNKYTAQ---FRDLNTKIGSRKII-----NRNQFN--YG 371
DB 318 FRFENKRYHIGGILERTQOTFTDMDTVPAFLTKAVFDANOKOAGSLRGKYGAGNHYG 377
QY 372 LSLNPYNLNLTAAYNSGRQKYPKSGFTGWLKGLKDFETYNNAKILDLNNTATFRLPRET 431
DB 378 -----GLFTSGENNAPVGAAY-GTGVFYD-----ET 402
QY 432 ELQTLTGFNFHNYGKNRPEELGLFFD--GPDQDNGL----- 468
DB 403 HTKSRYGLEYVYTNADKTDWADYARLSYDROGIGLDNHFQOHCSDGSDKYCRPSADKP 462
QY 469 YSYLGRKGDGLLPQKSTIVQAGSOYFNTFFDAALKDDIYRLNYSNTVGY-REGG- 526
DB 463 FSY---YKSDRVYIGESHKLLQAAFKKSFDTAKI-----RHNLSVN-LGYDREGSN 509
QY 527 ---EYTGYYGSDDEKFRAGFENSP----- 547
DB 510 LRHODY--YYQSAN---RAYSLKTPPONNGKKTSPNGREKNPWWISIGRNVVTRQICLF 564
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YKKRANNSHS-----I 579
DB 565 GNNTYDCTPRSINGKSYAAVRNVRNLGRWADVAGLRYDYRSTHSDGSGVSTGTHRTL 624
QY 580 SADGCDYFMPFASVSRTH-----RMPNIOEMYPFQSDGSGVHTALKPERANTWQFG--- 630
DB 625 SWNAGIVLKPADWLIDLTRYSTGRFLPSFAEMYGWRSGDKIKAVKIDPEKSFNKEAGIVF 684
QY 631 -----FNYYKGLLKQDDTLGLKLGVYRSRDN-----YIH----- 661
DB 685 KGDFGNLEASWNNAYRDLIVR-----GYEAQIKGCKEQVKGNPYNLAQASARITG 735
QY 662 -NHYGK--W---WDLNGDIP--SWYSSTGLAYT-IOHRNFKDKVHKHGFELNVDYGRFF 713
DB 736 INILGKIDWGVND---KLPEGWYST--FAYNRVRVEDIKKRDRTDIQSHL-FD----- 784
QY 714 TNL5AYQKSTQPTNFSDESSEPNNAKEDQDKOGYGLSRVSALPRDYGRLEVGTRWLG 773

DB 785 -----AIQPSRYV-----VSGGYDQPEGKWGVNGMLT-----YSKAKEITELLS 824
QY 774 KLTGGAMRYFGKSIRATAEERYIDGTNGGNTSNFRQLGKRSIKQETTLARQPLIFDFA 833
DB 825 RALLNG-----NSRNTKATARRTRPWYIVDVSG 852
QY 834 AYEPPKKNLIFRAEVKNLFDRIYI 856
DB 853 YYTVKKHFTLRAGVYNLLNRYV 875

RESULT 11
US-08-487-890A-94
Sequence 94, Application US/08487890A
Patent No. 5708149
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,890A
FILING DATE: 07-JUN-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-487-890A-94

Query Match 5.5%; Score 269; DB 1; Length 908;
Best Local Similarity 21.2%; Pred. No. 1.7e-13;
Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;

QY 5 FRLKPICFYLMGVLYHYSAEDAGRAGSEAIQVLEDVHVAK--RVPKDKKVFDTDARA 62
DB 7 FRLNILCLSLMTALPV---YAENV--QAEQAQEKQLDTIQVAKAKKQTRRDNEVTGLGKL 61

```
QY 63 VSTRQDIFKSSE-NLDNIVRSIPG-AFTQODK--SSGIYSLNIRGDSGFGFVNTWVDGIT 118
Db 62 VKSSDLSKQVNLNLDLTRYDPIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDDGVS 117
QY 119 Q-TFYSTSTDAG--RAGGSSQFSGVSDSNFIAGLDVVVYKFSFGSAGINSIAGSANSRLTLG 175
Db 118 QIOSYTAQAALGGTRTAGSSGAINIEYENKAVEISKGSNSSEYNGALAGSVAFOTKT 177
QY 176 VDDVVOGNTTYGLLLK-GLTGNTSTKGNMAAIGARKWLESASGVLYIGHRSRVAQNY 234
Db 178 AADIIEGKGWQIQSKTAYSGKDHALTQSLALAGR---SGGAELIITYTKRGREITHAH 233
QY 235 RVGGGQHIGNFECAEYLERKQ-----RYFVOEGAL-----KFNSD-SGKWER-- 276
Db 234 KDAGKG--VQSFNRLVDEDDKGGGQYRYFIVEECHNGYAAACKNKLKEDASVKDERKT 291
QY 277 -----DLQROQWKYKP-----YKNYNQELQKYIEGHDKSWRENLAPO 314
Db 292 VSTQDYTGNSRLNLANPLEYGSQSWLFRPGWHLDNRHVGVAVLERTQOTFTD--RDMTTPA 349
QY 315 YDTPTD--PSSLK-----QOSAGNLFKLEY-DGVFNKYTAQFRDLNNTK 355
Db 350 Y-FTSEDYVPSGLKGLGKYSGDNKAERLFVQEGGSTLQIGIGYGTGVF-----YDERHTK 402
QY 356 IGSRKIIINRYQNYGLSLNPTNLNLTAAYNSGRQKYPKGSFTGWGLLKDEFTYNNAK 415
Db 403 -----NR-----YGVVEY-VYHNADKDTWADYARLSYDRQG----- 431
QY 416 ILDNNTATPRLPRETELQTLTGFNTFHNHEYKGNRPPEELGLFFDGDQDNGLYSYLGRF 475
Db 432 -IDLN-----RLQOT---HCSHDSGNCRP-----DG---NKPYSP---Y 463
QY 476 KGBKGLLPQKSTIVQAGSYFNT-----FYFDAAL----- 506
Db 464 KSDRMIVESRNLFOAVFKAFDTAKIRHNLSINLGYDREKLSHSDYXLQNAVOAYDL 523
QY 507 -----KKDIYRLNYSNTVG---YRFGGE-YT-----GYGSDDEEK 539
Db 524 ITPKKPPFPNGSKNDYRVSIGTWTNTSPICRFGNNTYDTCPRNIGGNGYAAVODNV 583
QY 540 RAFGENSPYKKHCNRCGIYEPVLKYYKKRANNHSVS-----ISADFGDYFMPFASY 593
Db 584 R-----LGRWADVAGIRVDYRSTHSEDSKSVSTGTHRNLSWNAAGVVLKPTWM 631
QY 594 SRTH-----RPMNIQBMYSQIGDSGVHTALKPERANTWQFG-----FNT 633
Db 632 DLTYRSTGRFLPSFAEMYGWRAGESLKTLDLKPESFNREAGIVFKGDFGNLEASYFNN 691
QY 634 YKGLLKQDDTLGLKLVGYSRIDN-----YIH-----NVYKG--WWDLNG 672
Db 692 AYRDLI-----AFGYETRTQNGQTSASGDPGYRNAQNAIAGINILGKIDWHGVWG 742
QY 673 DIPSWSVSSTGLAY-----TIQHRNFKDKVHKHGFELNLDY--GRFFTNL 716
Db 743 GLPDGLYST-LAYNRKIKVDADIRADRTFTYSYLFDAVQSPRYVLGLGYDHPDGIWGINT 801
QY 717 SYAYQKSTQPTNFSASESPNNASKEDQLKQGYCLSRVSALPRDYGRLEVGTWGLNKL 776
Db 802 MFTYSK-----AKSVDEL-----LGSQAL 820
QY 777 LGGAMRYFGKSIRATAERYIDCTNGGNTSNFRLGKRSIKQETLARQPLIFDFYAAE 836
Db 821 LNG-----NANAKAASRR-----TRPWYVTDVSGYIN 848
QY 837 PKKNLIFRAEVKNLFDYRYI 856
Db 849 IKKHLTLRAGVYNLLNRYV 868
```

RESULT 12

US-08-478-435-94

; Sequence 94, Application US/08478435

; Patent No. 5922323

```
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,435
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-435-94
```

Query Match 5.5%; Score 269; DB 2; Length 908;

Best Local Similarity 21.2%; Pred. No. 1.7e-13;

Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;

QY 5 FRKPICFYLMGVLYHYHYAEDAGRAGSEAOIQVLEDDVHVAK--RVPKDKKVFYTDARA 62

Db 7 FRNLILCSLMTALPV---YAENV--QAEQAQEKQLDTIQVAKKAKQKTRDRNEVTGLGKL 61

QY 63 VSTRQDIFKSSE-NLDNIVRSIPG-AFTQODK--SSGIYSLNIRGDSGFGFVNTWVDGIT 118

Db 62 VKSSDLSKQVNLNLDLTRYDPIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDDGVS 117

QY 119 Q-TFYSTSTDAG--RAGGSSQFSGVSDSNFIAGLDVVVYKFSFGSAGINSIAGSANSRLTLG 175

Db 118 QIOSYTAQAALGGTRTAGSSGAINIEYENKAVEISKGSNSSEYNGALAGSVAFOTKT 177

QY 176 VDDVVOGNTTYGLLLK-GLTGNTSTKGNMAAIGARKWLESASGVLYIGHRSRVAQNY 234

Db 178 AADIIEGKGWQIQSKTAYSGKDHALTQSLALAGR---SGGAELIITYTKRGREITHAH 233

Db 464 KSDRMIEESRNLFOAVFKAFDTAKIRHNLINLGYDRFKSLSHSDYILQNAVQAYDL 523
QY 507 -----KKDIYRLNYSNTWV-----YRFGGE-YT-----GYIGSDDEFK 539
Db 524 ITPKKPPFPNGSKDNYPYVSGTGTNTSPICRFGNNTYDCTPRNIGGNGYAAVQDNV 583
QY 540 RAGENSPTKKHCNRCGIIYEPVLKYYKKRANHSVS-----ISADPGDFMPFPASY 593
Db 584 R-----LGRWADYGAGIRYDYRSTHSEDKSVSTGTHRNLSWNAVGVLPKPTWM 631
QY 594 SRTH-----RMPNIOEMFYSQIDSGVHTALKPERANTWQFG-----FNT 633
Db 632 DLYRSTGRFLSPFEMYCWGRAGESLKTLDLAPKSFNREAGIVFKGDFGNLEASYFNN 691
QY 634 YKGLLKQDDTLGLKLVGYSRIDN-----YH-----NYGK--WMDLNG 672
Db 692 AYRDLI-----AFGYETRTONGOTSASGDPGYRNAQARIAGINILGKIDWHGVWG 742
QY 673 DIPSWYSTGLAY-----TIOHRNFKDKVHKHGFELNLYDY--GRFFTNL 716
Db 743 GLPDGLYST-LAYNRKVKDADIRARTEFTVSYLFDVAVQPSRYVLGLGYDHPDGIWINT 801
QY 717 SYAYQKSTQPTNFSASESPNNASKEDOLKOGYGLSRVSALPRDYGRLVGTWRWLNKLT 776
Db 802 MFTYSK-----AKSVDEL-----LGSQAL 820
QY 777 LGGAMRYFGKSIRATAERYIDTNGTNSFNROLGKRSIKOTETLARQPLIFDYAAYE 836
Db 821 LNG-----NANAKKAASRR-----TRPWYVTDVSGYYN 848
QY 837 PKNLIFRAEVKNLFDPRYI 856
Db 849 IKKHLTLRAGVYNLLNRYV 868

RESULT 14

US-08-478-373-94
; Sequence 94, Application US/08478373
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murlin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-373-94

Query Match 5.5%; Score 269; DB.2; Length 908;
Best Local Similarity 21.2%; Pred. No. 1.7e-13;
Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;

QY 5 FRLKPICFYLMGVTLYHYSYAEDAGRAGSEAOIQVLEDVHVKAK--RVPKDKKVFTDARA 62
Db 7 FRLNLCLSLMTALPV---YAENV--QAEQAQERQDLTIQVAKKOKKTRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE--NLDNIVRSIPG-AFTQDK--SSGIVSLNIRGDSGFCGRVNTMVDGIT 118
Db 62 VKSSDLSLKEQVLNIRDLTRYDPGIAVVEQGRGASSG---YSIRG-MDKRVSLTVDGVS 117
QY 119 Q-TFYSTSTDAG--RAGGSSQFASVDNSFIAGLDVVKVYVSGFSAGNSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGGTTRTAGSSGAINIEYENVKAVEISKSNSSEYNGALAGSVAFOGKT 177
QY 176 VDDVVOGNTYGLLLK--GLTGTNSTKGNMAATCARWLESAGSVGLYGHSSRSVAQNY 234
Db 178 AADLIGEGKQWGIOSKTAYSGKDHALTQSLAGR-----SGGAELDIYTKRGRRETAH 233
QY 235 RVGGGQGHIGFNGAEYLERKQ-----RYFVQEGAL-----KFNSD-SGKWER-- 276
Db 234 KDAGKG--VQSFNRLVLEDEKKEGGSQYRYFIVEECHNGYAAACKNLKEDASVKDERKT 291
QY 277 -----DLORQWQKYP-----YKNNQELQKYLEGHDKSKRWENLAPQ 314
Db 292 VSTQDYTGSRNLLANPLEYGSQSWLFRPGWHLNHRHYGVAVLERTQOTFTD--RDMTPVA 349
QY 315 YDITPID--PSSLK-----QQSAGNLFKLEY-DGVFNKYTAQFRDLNTK 355
Db 350 Y-FTSEDYVPGSLKGLGKYSKDNKAERLFOGEGSTLQIGYGTGVF-----YDERHTK 402
QY 356 IGSRKIIINRNYPNYGLSLNPTNLTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAK 415
Db 403 -----NR-----YGVVEY-VYHNADKDTWADYARLSYDRQG----- 431
QY 416 ILDLNNTATFRLPRETELQTLGFNFHNEYGNKRNFEELGLFDPGDQDNGLYSYLGRF 475
Db 432 -IDLN-----RLQOT---HCHSDGSKNCRP-----DG-----NKPYSF---Y 463
QY 476 KDGKLLLPQKSTIVQAGSOYFNT-----FYFDAAL----- 506
Db 464 KSDRMIEESRNLFOAVFKAFDTAKIRHNLINLGYDRFKSLSHSDYILQNAVQAYDL 523
QY 507 -----KKDIYRLNYSNTWV-----YRFGGE-YT-----GYIGSDDEFK 539
Db 524 ITPKKPPFPNGSKDNYPYVSGTGTNTSPICRFGNNTYDCTPRNIGGNGYAAVQDNV 583
QY 540 RAGENSPTKKHCNRCGIIYEPVLKYYKKRANHSVS-----ISADPGDFMPFPASY 593
Db 584 R-----LGRWADYGAGIRYDYRSTHSEDKSVSTGTHRNLSWNAVGVLPKPTWM 631
QY 594 SRTH-----RMPNIOEMFYSQIDSGVHTALKPERANTWQFG-----FNT 633

Db 632 DLTYRSTGTFRLPSFAEMYGWRAGESLTKLTLKPEKSFNREAGIVFKGDFGNLEASYFNN 691
QY 634 YKGLLKQDDTLGLKLVGYRSRDN-----YIH-----NVYK--WMDLNG 672
Db 692 AYRDLI-----AFGYETRTQNGTOSASGDPGYRNAQNARIAGINILGKIDWHGCVG 742
QY 673 DIPSWSVSTGLAY-----TIQHRNFKDKVHKHKGFELELNVDY--GRFFTNL 716
Db 743 GLPDGLYST-LAYNRKVKDADIRADRTFTVSYLFDVAVQPSRYVLGLGYDHPDGIWGIN 801
QY 717 SYAYQKSTQPTNFSDAESPNNASKEDQKQGYCLRSVLSALPRDYGRLEVTGTRWLNKLT 776
Db 802 MFTYSK-----AKSVDEL-----LGSQAL 820
QY 777 LGGAMRYFGKSIRATAEERYIDGTNGTNTSNFROLGRKSIRKQETLARQPLIFDYAAE 836
Db 821 LNG-----NANAKAASRR-----TRPWYVTDVSGYYN 848
QY 837 PKKNLIFRAEVKNLFDRIY 856
Db 849 IKKHLTLRAGVYNLLNRYV 868

RESULT 15
US-08-474-671-94
; Sequence 94, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Van-Ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-474-671-94

Query Match 5.5%; Score 269; DB 3; Length 908;
Best Local Similarity 21.2%; Pred. No. 1.7e-13;
Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;

QY 5 FRUKPICFYLMGVTLHYHYSAEDAGRAGSEAQIOVLVDVHVAK--RVPKDKKVFTDARA 62
Db 7 FRNLILCLSLMTALPV---YAENV--QAEQAQEKOLDTIQVAKAKQKTRKRDNEVTGLGKL 61
QY 63 VSTRQDIFKSS--NLNIVRSIPG-AFTODK--SSGIVSLNIRGDSGFGFRVMTVDGIT 118
Db 62 VKSSDTLSKQVILNIRDLTRYDFGIAVVEOGRGASSG---YSIRG-MDKNRVSLTVDGVS 117
QY 119 Q-TFYSTSTDAG--RAGGSQFQASVDSNFIAGLDVVVKFSGFSAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGCTRTAGSSGAINIEYENVKAVEISKGSNSSEYNGALAGSVAFTQTKT 177
QY 176 VDDVQGNNTYGLLLK-GLTGTNSTKGNMAAIGARKWLESAGSVGLYGHSSRSVAQNY 234
Db 178 AADIIGEGKQWGIQSKTAYSGKDHALTQSLALAGR---SGGAELIILYTKRGRREIHAH 233
QY 235 RVGGGGOHIGNFGAEYLERKQ-----RYFVOEGAL-----KFNSD-SGKWER-- 276
Db 234 KDAGK--VOSEFNLVDEDDKKEGSGOYRFTVEEHCNGNYAACKNKLKEDASVKDERKT 291
QY 277 -----DLQRQWKYKP-----YKNYNNQELQYIEGHDKSWRENLAPO 314
Db 292 VSTQDYTGSNRLLANPLEYGSQSLFRPGWHLNDRHYVGAVLERTQQTFTD--RDMTVPA 349
QY 315 YDITPID--PSSLK-----QQSAGNLFKLEY-DGVFNKYTAQFRDLNNTK 355
Db 350 Y-FTSEDIYVPSLKGKLYSGDNKAERLFVQEGESTLQIGIGYGVF-----YDERHTK 402
QY 356 IGSRKIIINRYQFNGLSLNPTNLTAAYNSGRQKPKGSKFTGWLGLKDFETYNNAK 415
Db 403 -----NR-----YGVY-VYHNADKDWADYARLSYDRQ----- 431
QY 416 ILDNNTATFRLPRETELQTLGTFNHYFHNKRNPEELGLFFDQDQNGLYSYLGRF 475
Db 432 -IDLN-----RLQOT---HCSHDGSDKNCRP-----DG-----NKPYSE--Y 463
QY 476 KGDGLLPOKSTIVQAGSQYFNT-----FYFDAAL----- 506
Db 464 KSDRMIEESRNLFOAVFKKAFDTAKIRHNLSINLGYDRFKSQLSHSDYILQNAVQAYDL 523
QY 507 -----KKDIYRLNYSNTVY---YRFGGE-VT-----GYYSQSDDEPK 539
Db 524 ITPKKPPFPNGSKDNPIRVSIGTKTNTVTSPICRFNNTYDCTPRNIGNGNYTAAVODNV 583
QY 540 RAFGENSPYKKHCNRSCTGIEYVPLKKYKKRANNHSV-----ISADFGDYFMPFASY 593
Db 584 R-----LGRWADVAGIRYDYRSTHSEDSKSVSTGTHRNLSWAGVVLKPPFTWM 631
QY 594 SRTH-----RMPNIQEMYFSQIGDSGVHTALKPERANTWQFG-----FNT 633
Db 632 DLTYRSTGTFRLPSFAEMYGWRAGESLTKLTLKPEKSFNREAGIVFKGDFGNLEASYFNN 691
QY 634 YKGLLKQDDTLGLKLVGYRSRDN-----YIH-----NVYK--WMDLNG 672
Db 692 AYRDLI-----AFGYETRTQNGTOSASGDPGYRNAQNARIAGINILGKIDWHGCVG 742
QY 673 DIPSWSVSTGLAY-----TIQHRNFKDKVHKHKGFELELNVDY--GRFFTNL 716
Db 743 GLPDGLYST-LAYNRKVKDADIRADRTFTVSYLFDVAVQPSRYVLGLGYDHPDGIWGIN 801
QY 717 SYAYQKSTQPTNFSDAESPNNASKEDQKQGYCLRSVLSALPRDYGRLEVTGTRWLNKLT 776

Db	802	MFYISK-----AKSVDEL-----LGSQAL	820
Qy	777	LGAMRYFGKSI RATAERYIDTNGGNTSNFQLCKRSIKOTETLARQPLFDYAAYE	836
Db	821	LNG-----NANAKKAASRR-----TRPWYVTVDSGYIN	848
Qy	837	PKKNLIFRAEVKNLFDORRYI	856
Db	849	IKKHLTIRAGVYLLNRYV	868

Search completed: July 24, 2002, 08:53:59
Job time: 531 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using 'sw model

Run on: July 24, 2002, 08:58:25 ; Search time 223.16 Seconds
(without alignments)
1452.653 Million cell updates/sec

Title: US-09-762-926-6
 Perfect score: 4894
 Sequence: 1 MRSSRLKPIGFYLMGVTL.....SVLTNFAAGRTFLITMSYKE 921

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 S

Database : Pending Patents AA Main:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4894	100.0	921	21	US-09-762-926-6
2	4779.5	97.7	922	21	US-09-762-926-4
3	4769.5	97.5	922	21	US-09-762-926-2
4	4707.5	96.2	922	17	US-09-303-518D-884
5	4703	96.1	888	17	US-09-303-518D-878
6	4696.5	96.0	922	17	US-09-303-518D-880
7	4348.5	88.9	871	17	US-09-303-518D-882
					Sequence 6, Appli
					Sequence 4, Appli
					Sequence 2, Appli
					Sequence 884, App
					Sequence 878, App
					Sequence 882, App
					Sequence 880, App

ALIGNMENTS

```

RESULT      1
US-09-762-926-6
, Sequence 6, Application US/09762926
, GENERAL INFORMATION:
, APPLICANT: Thonnard, Joelle
, TITLE OF INVENTION: Novel Compounds
, FILE REFERENCE: BM45330
, CURRENT APPLICATION NUMBER: US/09/762,926
, CURRENT FILING DATE: 2001-02-14
, PRIOR APPLICATION NUMBER: PCT/EP99/05989
, PRIOR FILING DATE: 1999-08-13
, PRIOR APPLICATION NUMBER: G5 9818004.5
, PRIOR FILING DATE: 1998-08-18
, NUMBER OF SEQ ID NOS: 9
, SOFTWARE: FastSeq for Windows Version 4.
, SEQ ID NO 6
, LENGTH: 921
, TYPE: PRT
, ORGANISM: Bacteria
US-09-762-926-6

```

	Query Match	100.0%;	Score 4894;	DB 21;	Length 921;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches ^921; Conservative	0;	Mismatches	0;	Gaps 0;
QY	1	MRSSFRLKPTCFYLMGVTLYHYSYAEDAGRAGSQAQVLEVDHVHKARVPKDKKVF	60		
db	1	MRSSFRLKPTCFYLMGVTLYHYSYAEDAGRAGSQAQVLEVDHVHKARVPKDKKVF	60		

Qy	61	RAVSTRQDIFKSSNNLNI	VRSPGATQODKSSGIVSLNIRGDSGFRVMTWDG	ITOT 120
Db	61	RAVSTRQDIFKSSNNLNI	VRSPGATQODKSSGIVSLNIRGDSGFRVMTWDG	ITOT 120
Qy	121	FYSTSDAGRAGGSQF	GCASVDNSNFTAGLDVWKGFSFGSAGINSLAGSANLRTL	GLGVDDVV 180
Db	121	FYSTSDAGRAGGSQF	GCASVDNSNFTAGLDVWKGFSFGSAGINSLAGSANLRTL	GLGVDDVV 180
Qy	181	QGNNTYGLLLKGLTGT	NSTKGNMAAATGARKWLESGASGVLYGHSSRRSVAQNRY	RVGGG 240
Db	181	QGNNTYGLLLKGLTGT	NSTKGNMAAATGARKWLESGASGVLYGHSSRRSVAQNRY	RVGGG 240
Qy	241	QHIGNFAEYLERRKQRY	FVQEGALFNDSGKWERDLQROQWKYPKYNNNOELOKYI	300
Db	241	QHIGNFAEYLERRKQRY	FVQEGALFNDSGKWERDLQROQWKYPKYNNNOELOKYI	300
Qy	301	EGHDKSWRENLAPOYD	ITPIDPSLKOOSAGNLFKLEYDGVFNKYTAQPRDLNTKIGSRK	360
Db	301	EGHDKSWRENLAPOYD	ITPIDPSLKOOSAGNLFKLEYDGVFNKYTAQPRDLNTKIGSRK	360
Qy	361	IINRNYOFNYGLSNPYN	TNLNLTAAYNSGROKYPKGSKTGWGLLKDFETYNNAKILDLN	420
Db	361	IINRNYOFNYGLSNPYN	TNLNLTAAYNSGROKYPKGSKTGWGLLKDFETYNNAKILDLN	420
Qy	421	NTATFRLPRETELOTT	LGPNFYHNEYGNKFRPEELGLFFDGDQDONGLYSLGRFKPGDKG	480
Db	421	NTATFRLPRETELOTT	LGPNFYHNEYGNKFRPEELGLFFDGDQDONGLYSLGRFKPGDKG	480
Qy	481	LLPKOSTIIVOPAGSQ	YENTFYFPAALKKDIIYRLNYSNTVGRFGEYTYGYSDDDEPKR	540
Db	481	LLPKOSTIIVOPAGSQ	YENTFYFPAALKKDIIYRLNYSNTVGRFGEYTYGYSDDDEPKR	540
Qy	541	AFGENSPTYKKHCNR	SGIYEPVLKKYKKRANHSHSVISADFQDFMPFASYSRTHRP	600
Db	541	AFGENSPTYKKHCNR	SGIYEPVLKKYKKRANHSHSVISADFQDFMPFASYSRTHRP	600
Qy	601	NIQWYFSQIGDSGVHT	ALPERANTWQGFNTYKKGLLKQDDTLGLKLUVGYRSRIDNYI	660
Db	601	NIQWYFSQIGDSGVHT	ALPERANTWQGFNTYKKGLLKQDDTLGLKLUVGYRSRIDNYI	660
Qy	661	HNVYKGMWDLNGDI	PSWVSTGLAYTIOHRNFKDKVHKHGFLELNYDYGREFTNL	SVAY 720
Db	661	HNVYKGMWDLNGDI	PSWVSTGLAYTIOHRNFKDKVHKHGFLELNYDYGREFTNL	SVAY 720
Qy	721	OKSTQPTNFS	SDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGT	RWLGNKLT LGGA 780
Db	721	OKSTQPTNFS	SDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGT	RWLGNKLT LGGA 780
Qy	781	MRYFGKSIRATAERY	IDGTNGTGNSTNFCOLGKRSIKQETLAROPLIFDFFAAYEPKKN	840
Db	781	MRYFGKSIRATAERY	IDGTNGTGNSTNFCOLGKRSIKQETLAROPLIFDFFAAYEPKKN	840
Qy	841	LIFRAEYKVNLFDRRY	IDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTL	CNGKYGCTS 900
Db	841	LIFRAEYKVNLFDRRY	IDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTL	CNGKYGCTS 900
Qy	901	KSVLTNFARGTFLIT	MSYKF 921	
Db	901	KSVLTNFARGTFLIT	MSYKF 921	

RESULT 2

US-09-762-926--4	
: Sequence 4, Application US/09762926	
: GENERAL INFORMATION:	
: APPLICANT: Thonnard, Joelle	
: TITLE OF INVENTION: Novel Compounds	
: FILE REFERENCE: BM45330	
: CURRENT APPLICATION NUMBER: US/09/762,926	
: CURRENT FILING DATE: 2001-02-14	
: PRIOR APPLICATION NUMBER: PCT/EP99/05989	
: PRIOR FILING DATE: 1999-08-13	
: PRIOR APPLICATION NUMBER: GB 9818004.5	

QY 900 SKSVLTNFARGRTFLITMSYKF 921
Db 901 SKSVLTNFARGRTFLITMSYKF 922
RESULT 3
US-09-762-926-2
; Sequence 2, Application US/09762926
; GENERAL INFORMATION:
; APPLICANT: Thonnard, Joelle
; TITLE OF INVENTION: Novel Compounds
; FILE REFERENCE: BM45330
; CURRENT APPLICATION NUMBER: US/09/762,926
; CURRENT FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: PCT/EP99/05989
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: GB 9818004.5
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Bacteria
US-09-762-926-2

Query Match 97.5%; Score 4769.5; DB 21; Length 922;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 901; Conservative 8; Mismatches 12; Indels 1; Gaps 1;
QY 1 MRSSFLKPCFCYLMGVLYHYSYAEDAGRAGSEAOIQVLEDDVHVAKRVPKDKKVFDTA 60
Db 1 MRSSFLKPCFCYLMGVLYHYSYAEDAGRAGSEAOIQVLEDDVHVAKRVPKDKKVFDTA 60
QY 61 RAVSTRODIFKSSNLDNIIVRSIPGAFTQDDKSSGIVSLNIRGDSGFRVNTWVDGITQT 120
Db 61 RAVSTRODIFKSSNLDNIIVRSIPGAFTQDDKSSGIVSLNIRGDSGFRVNTWVDGITQT 120
QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASVGLYGHSSRRSVAQNYRVGGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASVGLYGHSSRRSVAQNYRVGGGG 240
QY 241 OHIGNFCAEYLERRKQRYFVQEGALKFNSDGKWERDLQROQWKYKPYKNYN-OELQKY 299
Db 241 OHIGNFCAEYLERRKQRYFVQEGALKFNSDGKWERDLQROQWKYKPYKNYN-OELQKY 300
QY 300 IEHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 359
Db 300 IEHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360
QY 360 KIINRNQFYNGSLNPTNLNTAAYNSGRQKYPKGSKFTGWLKGLDKFTYNNAKILDL 419
Db 360 KIINRNQFYNGSLNPTNLNTAAYNSGRQKYPKGSKFTGWLKGLDKFTYNNAKILDL 420
QY 420 NNTATFLPRETELQTTLGFNHYNEHVGKRPPEELGLFPDGDQDNGLSYLGREFKGD 479
Db 420 NNTATFLPRETELQTTLGFNHYNEHVGKRPPEELGLFPDGDQDNGLSYLGREFKGD 480
QY 480 GLLPQKSTIVQAGSOYFNTFYDAAALKDIIYRLNSTNTVGRFGGEYTYGYSDDDEFK 539
Db 480 GLLPQKSTIVQAGSOYFNTFYDAAALKDIIYRLNSTNTVGRFGGEYTYGYSDDDEFK 540
QY 540 RAFGENSPYKKHCNRCISGIEPVLYKYGKKRANNHVSISADFGDYFMPFASYSRTHRM 599
Db 540 RAFGENSPYKKHCNRCISGIEPVLYKYGKKRANNHVSISADFGDYFMPFASYSRTHRM 600
QY 600 PNIOEMVFSQIGDSGVITALKPERANTWQGFNTYKGLLKQDDTLGLKLVGYRSRDNY 659

Db 601 PNIOEMVFSQIGDSGVITALKPERANTWQGFNTYKGLLKQDDTLGLKLVGYRSRDNY 660
QY 660 IHNVYGKWDLNGDIPSWVSSTGLAYTIQHRNFKDKVHKHGFLELANYDYGRFFTNLSYA 719
Db 661 IHNVYGKWDLNGDIPSWVSSTGLAYTIQHRNFKDKVHKHGFLELANYDYGRFFTNLSYA 720
QY 720 YOKSTQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWGLNKLTLGG 779
Db 721 YOKSTQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWGLNKLTLGG 780
QY 780 AMRYFGKSIIRATAERYIDGTNGGNTSNFROLGKRSTKQETTLARQPLIFDYAAYEPKK 839
Db 781 AMRYFGKSIIRATAERYIDGTNGGNTSNFROLGKRSTKQETTLARQPLIFDYAAYEPKK 840
QY 840 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEEDVTCNADKTLCKNGKYGT 899
Db 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEEDVTCNADKTLCKNGKYGT 900
QY 900 SKSVLTNFARGRTFLITMSYKF 921
Db 901 SKSVLTNFARGRTFLITMSYKF 922
RESULT 4
US-09-303-518D-884
; Sequence 884, Application US/09303518D
; GENERAL INFORMATION:
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; TITLE OF INVENTION: Neisserial Antigens
; FILE REFERENCE: CHIR0160
; CURRENT APPLICATION NUMBER: US/09/303,518D
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 884
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-09-303-518D-884

Query Match 96.2%; Score 4707.5; DB 17; Length 922;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 885; Conservative 15; Mismatches 21; Indels 1; Gaps 1;
QY 1 MRSSFLKPCFCYLMGVLYHYSYAEDAGRAGSEAOIQVLEDDVHVAKRVPKDKKVFDTA 60
Db 1 MRSSFLKPCFCYLMGVLYHYSYAEDAGRAGSEAOIQVLEDDVHVAKRVPKDKKVFDTA 60
QY 61 RAVSTRODIFKSSNLDNIIVRSIPGAFTQDDKSSGIVSLNIRGDSGFRVNTWVDGITQT 120
Db 61 RAVSTRODIFKSSNLDNIIVRSIPGAFTQDDKSSGIVSLNIRGDSGFRVNTWVDGITQT 120
QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASVGLYGHSSRRSVAQNYRVGGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASVGLYGHSSRRSVAQNYRVGGGG 240
QY 241 OHIGNFCAEYLERRKQRYFVQEGALKFNSDGKWERDLQROQWKYKPYKNYN-OELQKY 299
Db 241 OHIGNFCAEYLERRKQRYFVQEGALKFNSDGKWERDLQROQWKYKPYKNYN-OELQKY 300
QY 300 IEHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 359
Db 301 IEHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360

; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 882
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-09-303-518D-882

```
Query Match      96.0%; Score 4696.5; DB 17; Length 922;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 883; Conservative 15; Mismatches 23; Indels 1; Gaps 1;

Qy 1 MRSSFLKPCFCYLMGVTLHHYSAEDAGRAGSEAOQVLEHVHAKRVPKDKKVFETDA 60
Db 1 MRSSFLKPCFCYLMGVTLHHYSAEDAGRAGSEAOQVLEHVHAKRVPKDKKVFETDA 60

Qy 61 RAVSTRODIFKSSBNLONIVRSIPGATFQDDKSSGIVSLNIRGDSGFRVNTWVDGITOT 120
Db 61 RAVSTRODVFKSGNENLONIVRSIPGATFQDDKSSGIVSLNIRGDSGFRVNTWVDGITOT 120

Qy 121 FYSTDAGRAGGSSQFCASVDSNFIAGLDVYVKGFSGSAGINSLAGSANLRTILGVDDVV 180
Db 121 FYSTDAGRAGGSSQFCASVDSNFIAGLDVYVKGFSGSAGINSLAGSANLRTILGVDDVV 180

Qy 181 QGNNTYGLLLKGLTGTNSTKGNMAAATGARKWLESASGVLYGHSSRVAQNTYRVGGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAATGARKWLESASGVLYGHSSRVAQNTYRVGGGG 240

Qy 241 QHIGNFGAELERKORYFVQEGALKFNDSQGWKWERDLQROOKYKPKYKNYN-QEIQKY 299
Db 241 QHIGNFGEELERKQQYFVQEGGLKFNAGSGKWERDLQRYWTKWYKYYEDPOELQKY 300

Qy 300 IEHDKSWENRAPQYDITPIDPSLQOOSAGNLFKLEYDGVFNKYTAQPRDLNTRIGSR 359
Db 301 IEHDKSWENRAPQYDITPIDPSGLQOOSAGNLLNLEYDGVFNKYTAQPRDLNTRIGSR 360

Qy 360 KIINRNTQFNGLSLNPTYNLNTAAVNSGRQKYPKSGKSTGWGLLKDFETYNNAKILDL 419
Db 361 KIINRNTQFNGLSLNPTYNLNTAAVNSGRQKYPKSGKSTGWGLLKDFETYNNAKILDL 420

Qy 420 NNTATFLPRETELQTTLGFNYFHNEYGKNRFPBELGLFDDGDDNGLYSLGRFPGDK 479
Db 421 NNTATFLPRETELQTTLGFNYFHNEYGKNRFPBELGLFDDGDDNGLYSLGRFPGDK 480

Qy 480 GLLPQKSTIVOPAGSQYFNTFYDAALKDIIYRLNYSNTNVTNVTNVTNVTNVTNVTNVTN 539
Db 481 GLLPQKSTIVOPAGSQYFNTFYDAALKDIIYRLNYSNTNVTNVTNVTNVTNVTNVTNVTN 540

Qy 540 RAFGENSPYKHKCNRSQGIYEPVLKKGKRRANNHNSVTSADPGDYFMPFPASVSRTHRM 599
Db 541 RAFGENSPAYKEHCDPSCGIEPVLLKKGKRRANNHNSVTSADPGDYFMPFPAGYSRTHRM 600

Qy 600 PNIQEMYFSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYRSRDNV 659
Db 601 PNIQEMYFSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYRSRDNV 660

Qy 660 IHNVYKQWDLNDIPSWGSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFFTNLSYA 719
Db 661 IHNVYKQWDLNDIPSWGSTGLAYTIHRNFKDKVHKHGFLELNYDYGRFFTNLSYA 720

Qy 720 YOKSTQPTNFSDESPPNASKEDQLKOGYGLSRVSALPRDYGRLEVTWGLNKLTLGG 779
Db 721 YOKSTQPTNFSDESPPNASKEDQLKOGYGLSRVSALPRDYGRLEVTWGLNKLTLGG 780

Qy 780 AMRYFGKSIATAEERYIDGTNGNTSNFQLGKRSIKQETTLARQPLIFDYAAAYEPKK 839
Db 781 AMRYFGKSIATAEERYIDGTNGNTSNVQLGKRSIKQETTLARQPLIFDYAAAYEPKK 840

Qy 840 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEDTVCNADKTLCKNGKYG 899
Db 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEDTVCNADKTLCKNGKYG 900
```

```
Qy 900 SKSVLTNPFARGRTTELITMSYKF 921
Db 901 SKSVLTNPFARGRTTELITMSYKF 922

RESULT 7
US-09-303-518D-880
; Sequence 880, Application US/09303518D
; GENERAL INFORMATION:
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisserial Antigens
; FILE REFERENCE: CHIR0160
; CURRENT APPLICATION NUMBER: US/09/303,518D
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 880
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (22)..(22)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (38)..(38)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (40)..(40)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (45)..(45)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (52)..(52)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (68)..(68)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (95)..(95)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (124)..(124)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (133)..(133)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (277)..(277)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (355)..(355)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (370)..(370)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (414)..(414)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (477)..(477)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (483)..(483)
; OTHER INFORMATION: Xaa= any amino acid
```

Query Match	52.9%	Score	2588.5	DB	21	Length	947
Best Local Similarity	55.6%	Pred. No.	4e-232				
Matches	503	Conservative	144	Mismatches	232	Indels	25
Gaps	12						
Qy	39	VLEDVHVYKAKR-VPKDKKVFYTDARAVSTRODIFKXSENLDNIVRSIFGAFYQODKSSGIV	97				
		: : : : : : : : : : : : : : : : : : : :					
Db	48	ILDEVVYATATNGTKTSQKPFYKASATSVRENVFNASENIDAIVRSVPGAFYQODKSSGLV	107				
Qy	98	SLNIRGDSGFGFRVNTWVDGITQTFYSTSTDRAGAGSGSOFCAVSDNSFIAGLDVVKVGSFS	157				
		: : : : : : : : : : : : : : : : : : : :					
Db	108	SLNVRGDSGFGFRVNTWVDGITQTFYSTSTDRAGGGTSGFCAVDQNFIAVGLNKVGSFN	167				
Qy	158	GSAGINSLAGSANLRTLGVDVVGQNTNYGLLLXGLTGTNSTKGNAMAAGARKWLESQA	217				
		: : : : : : : : : : : : : : : : : : : :					
Db	168	GKGGINTLTGSANFTLNADDYIKDKKNFGFIAGLTKNATDKNEMLAGRGGLDNG	226				
Qy	218	SVGLVYGHRSRVAQNYRVGGGHHGNGFAEYLERRKQRYFVQEGALKNFSDSGKWEDR	277				
		: : : : : : : : : : : : : : : : : : : :					
Db	227	SISALYAYSHKDISONYKVGGGGTHIGNVGDLLLSKQKQVFAKEHALTYNEASRSWQKD	286				
Qy	278	LQROO-----W--KYK-PYKNY-----NNQELQKYEIIEGHDKSWENLAPDYITPID	321				
		: : : : : : : : : : : : : : : : : : : :					
Db	287	LTKLDKETGKPLWRKYQFGGKGLGICIDTFKEFDEVDVADQOQWQKHGAKEYSITPID	346				
Qy	322	PSSLKQOSAGNLFLEYDGVFNKYTAQFRDLNTKIGSRKIIINRYQFNYGSLSLNPNYNLN	381				
		: : : : : : : : : : : : : : : : : : : :					
Db	347	ITALNQTSKSHAKIRYNDTSDVGLQRKMDWTIGSRISNDNYQLDAAYNPNEIDLK	406				
Qy	382	LTAAYNSGRQYKPGSKFTGMGLLKDFETYNNAKILDLNNTATFRLPRELTOQLTFGNY	441				
		: : : : : : : : : : : : : : : : : : : :					

Db 957 -----DV-----LAFILGRGRTPLOGLTLEVOF 977

RESULT 12

US-09-252-991A-27095
; Sequence 27095, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: 1999-02-18
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27095
; LENGTH: 992
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27095

Query Match 8.7%; Score 426; DB 16; Length 992;
Best Local Similarity 22.9%; Pred. No. 1.5e-29;
Matches 235; Conservative 132; Mismatches 377; Indels 284; Gaps 44;

QY 22 YSAEDAGRAGSEAOIQV-----LEDVHVKAKRV-PKDKKVF 58
Db 121 YSVLEGLQRLVSLTGLQVAGTSGAYLLVPRMPGETPADLSPVYSAELADPQRETYT 180
QY 59 DARAVSTRQDIFKSSNLD-----NIVRSIPGFTQODKSSGIVSLNIRGDSGFRV 110
Db 181 -----APRSSVYLSSSIDRFRVRSVGDLLQIGPGVQSGDRGGALDVNIRGQGSRV 235
QY 111 NTWVDGITOTFYSTSDAGRAGSSQFGASVDSNFAGLDVVKGSFGSGAGINSAGSAN 170
Db 236 AVRVDGAEQ-----ALDVRYGAGTQORSYIDPDLVSSVTVDKGPSTRSGAI---GGSVE 287
QY 171 LRLTGVDVVQGNNTYGLLLKLTGNTSGNMAAATGARKWLESGASV-----GVLYGH 225
Db 288 MRTIGVKDILVDGKDLGVRGTGDVWNN--GVAPQHRSAASKTENLSSVPHDRGSLFGS 344
QY 226 SRR--SVAQNYRVGGGQHGTFGAELYERRKORYFVQEGALKFNDSGKWERDLORQW 283
Db 345 QAKSGSAFAFYR---NEHL-DLVAAYAOQNGNYF-----SGKKGD----- 382
QY 284 KYRKYNNQOE---LQKYIEGHD-----KSWRENLAPOYDIT----- 318
Db 383 RYRYNNRYGSESSVAKVYNAGLEVLSNSETESYLLKATWR--IADEHTLDLGYRYYDG 440
QY 319 ---PIDPSSLKQOSAGNLFLEYDGV-FNKYTAQFRLNTKIGSRKIINRYQFNGLSL 374
Db 441 RTGEIMPSDIFRGTAGIYQIPSEWKIDITYTARYLPEN-----NPLVDLSTGLMW 493
QY 375 NPYTNLNLTAAYNSGRQYKPGSKFTGWLLKDFETYNNNAKI-LDLNNTATFRL----- 427
Db 494 TDAKSDMLTVLAPRSQAYSRRNWT-----RODNERRIGDNLNVARFETDFGDFK 544
QY 428 -----PRETELQTTIGFN-----YFNHEYKKN-----RPFEEGLGLFDG- 461
Db 545 LDLGGSFQVEDIQPKQSVVTLHDINANRTLDRATROEYGLNGKLEFKPVERLTLWGGR 604
QY 462 ----POQDNGLYSLGRFGDKLLPKOKSTIVOPA--GSOYF---NTFFYDAA---LKK 508
Db 605 YSHFNKDNIGISAPRREDMDRFI-----TVSRPGYGYMMWFPDQNGQYTDATDPLNN 660
QY 509 DIYRLNLTSTNVGRFGGEYTYGSDDEPKRAFGENSPYKKHCNRCSCGIYEPV-LKKY 567
Db 661 GIVNTNNTNPFEGIPF-----DEFGPA-----NVTVHPSRVTVNTVGYNYSKK 703

QY 568 GKRRANNHVSISADF---GDYFMPFASYSRTHMPNIOEMYSQIGDSGVHTA--LKPE 622
Db 704 GSRGGGSPAFGINFELAPDTFV-YASYTEGLRLPSLFE---TSQGTLOVEPGDKLKE 759
QY 623 RANTWQFGFNTYKKGLLKQODTLGLKLVGYRSRIDNVIHNVYKWKWDLNGLDIPSWSSG 682
Db 760 RSRWEIGASALRDSLLADGDSAAIKLAYFNNTKNTYTRYD-----PQMGMLT 810
QY 683 LAVTIOHRNFKDKVHKHGFLELNVDYGRFTNLSYAYQKSTQPTNFSASESPNNASKE 742
Db 811 FSNT-----DSVRTSGLEQSHYDAGRFAVADLSATYILKTTCTCDAFAARLAGANRY 863
QY 743 DOLKQGVLSRVSAL-----PRDYGLEVGTWLNKLTGLGAMRYEGKIRATAE 793
Db 864 QRTENTPNCPTGSGFMGYTNTQPPRLATNLTAGLRFDOALTLLGGRWY--TSQPTATAD 922
QY 794 ERYIDGTNGGNTSNFROLGRRSIKQETELARQPLIFDYAAEYKKNLIFRAEYKNLFR 853
Db 923 KPWQVG-----ATTPQIEYRSVQ-----LPDLFLKYLFEHTELNASLQNLTD 966
QY 854 RYIDPLDAGNDAATQRYYSFEDPKDKEDVTCNADKTLGNGKYGGTSKSVLTNFAGRTF 913
Db 967 YYLDPL-----AQSFMP-----APGRTL 984
QY 914 LITMSYKF 921
Db 985 RVGMQAKF 992

RESULT 13
US-09-897-516-4927
; Sequence 4927, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 4927
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-4927

Query Match 8.4%; Score 410.5; DB 22; Length 364;
Best Local Similarity 29.5%; Pred. No. 8.1e-29;
Matches 113; Conservative 54; Mismatches 125; Indels 91; Gaps 12;

QY 74 ENLDNIVRSIPGFTQODKSSGIVSLNIRGDSGFRVNTWVDGITOTFYSTSDAGRAG 133
Db 39 ESMDSVLRSPGTYTQMDTSGTIANIRMGSGFRVNMVMDVGSQSFYIAPSEFAHGT 98
QY 134 S--SQFASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGLVDVVOGNNNTYGLLK 191
Db 99 QPNYOTGALIDSFIIRTDVDRGOANDSDSNALVGSANFTIGDIDVIEGNKLGILTK 158
QY 192 GLGTNTSKGNAMAAGAR-KWLESASGVLYGHRSRVAQNYRVGGGQHGIGNFAEY 250
Db 159 SAYGTNGLGKNGMIAIAGRTQAFSTEGSIGAMLAISGHSIDAHYKNAM----- 207
QY 251 LERRKQRYFVQEGALKFNDSGKWERDLORQWQKYPKYKNYNNQELQKYLEGHDKSHREN 310

Db 208 -----VSSEEGTD-----KTFNREPNSQLMKI-----N 231
QY 311 LAPQYDITPIDPSSLKQOSAGNLFKLEVDGVE--NKYTAQFRDLNTKIGSRKIINRYQF 368
Db 232 IKP-----NDPHELSGRFYHNKT-----KRHIDSVDYIL 263
QY 369 NYGLSLNPYTNL-----NLTAAYNSGRQKYPKSGKFTGWLKDKDFETYNNAKILDNLNTATF 425
Db 264 KY--HYTPFSELIDTNILLGSGKGNQYFVKMSGLGKG-----ESHKNSNTIDIKNTSRF 316
QY 426 RLPRETELQTLGFNYFHNEXGK 448
Db 317 NY-GETDISFTLGSKLMDEYHK 338
RESULT 14
US-60-215-161-4927
; Sequence 4927, Application US/60215161
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)A
; CURRENT APPLICATION NUMBER: US/60/215,161
; CURRENT FILING DATE: 2000-06-30
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 4927
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-60-215-161-4927
Query Match 8.4%; Score 410.5; DB 26; Length 364;
Best Local Similarity 29.5%; Pred. No. 8,1e-29;
Matches 113; Conservative 54; Mismatches 125; Indels 91; Gaps 12;
QY 74 ENLDNIVRSIPGATQODKSSGIVSLNIRGDSGFRVNTMVDGITQFYSTSTDAGRAGG 133
Db 39 ESMDSVLSPLPGTYTQMDTSQGTIAVNRGMSGFRVNMVGVQSQSFYIAPSEFAHG 98
QY 134 S--SOFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVVOGNNTYGLLLK 131
Db 99 QPNYQTGALIDSFIITDVRGOANDSDSVNALVGSANFRTIGIDDVIFEGNKLGLTK 158
QY 192 GLTGTNSTKGNMAAIGAR-KWLESAGSVGVLYGHSRRSVAQNYRVGGGQGHGNGAEY 250
Db 159 SAYGTNGLKNGMIATAGRTQAFSTEGSIGAMLAISGSHIDAHKNAG----- 207
QY 251 LERRKORYFVOEGALFNSDGKWERDLQROQWKYPKYNNOELQKYTEGHDKSWREN 310
Db 208 -----VSSEEGTD-----KTFNREPNSQLMKI-----N 231
QY 311 LAPQYDITPIDPSSLKQOSAGNLFKLEVDGVE--NKYTAQFRDLNTKIGSRKIINRYQF 368
Db 232 IKP-----NDPHELSGRFYHNKT-----KRHIDSVDYIL 263
QY 369 NYGLSLNPYTNL-----NLTAAYNSGRQKYPKSGKFTGWLKDKDFETYNNAKILDNLNTATF 425
Db 264 KY--HYTPFSELIDTNILLGSGKGNQYFVKMSGLGKG-----ESHKNSNTIDIKNTSRF 316
QY 426 RLPRETELQTLGFNYFHNEXGK 448
Db 317 NY-GETDISFTLGSKLMDEYHK 338
RESULT 15

US-09-897-516-6415
; Sequence 6415, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 6415
; LENGTH: 937
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-6415
Query Match 7.7%; Score 376.5; DB 22; Length 937;
Best Local Similarity 19.5%; Pred. No. 6e-25;
Matches 212; Conservative 150; Mismatches 380; Indels 343; Gaps 40;
QY 27 DAGRAGSEAOI-OVLEDVHVHAKRVPKDKKVFYTDARAVSTRQDIFKSSSEMLNIVRSIPG 85
Db 6 DGEAGYDAVYDKDISNIYIKKEIER-----YKGASPAD-VIRGAVG 47
QY 86 AFTQODKSSGIVSLNIRGDSGFRVNTMVDGITQFYSTSTDAGRAGGSQFGASVDSNF 145
Db 48 VYSGDARNSGALDINIRGQGGRIPIVTIDGTEQ-----AITVGRYNGANNRYIDPNL 102
QY 146 IAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVVOGNNTYGLLLKGLTGTNSTK----- 200
Db 103 ISSIEIEKGPSLNRVRKVGSGGAVSIKTLNIDVDVPEGETFGINTKLETSSNSVKERTPS 162
QY 201 -----GNAMAAIGARKWLESAGSVGVLYGHSRRSVAQNYRVG-CGGQHIGN 245
Db 163 LSLGQDYRDVNPNIHNGIESDPALKITPHSSKDNKLFKFDNA---FRVAVGTROEYFD 218
QY 246 FGAEYLERKQRYFVOEGA-----LKEN----- 268
Db 219 LMLAYAYRHKNYFACKGGAHRYDAAATEADINLMLNPKTSLDPLPFAARIYRPGNEVP 278
QY 269 SDGSKWERDLQROQWKYPKYNNOELQKYTEGHDKSW-----BENLAPOY 315
Db 279 NTSSKMQSVLKNTHFT-----DEQALQAFNRTRMEFGDIMPSRLASVLAKENSVPQW 333
QY 316 DITPIDPSSLKQOSAGNLFKLEVDGVFNKY-----TAQFRDLNTKIG-SRKLIINRN 365
Db 334 PL-----ANARQOAAASLNLYK--WDSANPYIDFDMNLWTTRTISNTSGGYPGRVTRD 386
QY 366 YQF--NYGLSLNPYTNLNTAAVNSGRQKYPKSGKFTG--LLKDFETYNNAKILDNLN 421
Db 387 YDWEGGKGRSTIDGTLINTAVTNAQNNR-----WGVDISNKFELTQNLDLTLMGN 437
QY 422 TATFRLPRETEL-----QTLGLFNY----- 441
Db 438 FQBERLGSDDDSHIDNLYFFQSPARKGQEQEINLAFNFDWRPTSWALGAGAKRYSYWS 497
QY 442 ----FHNE--YGNRPFPEELGLFF-----DGPD----QDNGLYSYIGR----- 474
Db 498 KODFLNERRIARDRNYKEPGEIIGKMSYWRVTVEDEANDFKRNDKEYYSKLSRTEKRE 557
QY 475 -----FKGDKGLLPQKSTIVO-----PAGSOYF-----NTFFYDAAL 506
Db 558 LRAKTKKIKDRNDRIQOKRWLVEEKFEWKYNPETHGKLNKSDNPYFNGLQDMNEKVIDPTS 617

```
Qy 507 KDIYRLNYSNTVGYRFGGEYTGYYGSDDEFKRAFGENSPYKKKHONRSCGIYEPVLKK 566
Db 618 GREAYKEYSANSL-----SNDKILUSQDVDP-----WEPAPKR 652
Qy 567 YGKRANNHSVSI SADFGDYFMPFASYSRTHRMPIOE--MYFSQIGDSGVHTALKPERA 624
Db 653 --KSHAWAPTFSATAYITDDLRVYTRYAEAVRMPISIFEDIVGFSGVKENYIGLKFPERA 710
Qy 625 NTWQFGFNTYKGLLKQDDTLGLKLVYRSRIDNYI--HNVYKMWDLNGDIPSWSSTG 682
Db 711 KTIETGFYDFSQLVNAERNADIKLSYNTVNIENVFDRDNTY----- 752
Qy 683 LAYTIQHRNEK--DKVHKHGFELNVDYGRFFTNLSYAYOKSTQPTNFSAS--ESPNN 738
Db 753 -----NFAQLDKQKLAGLEQARYDNGSFFDMLGLVNMKNKVCNNSAARMDSQNR 804
Qy 739 ASKEDQLKQY--GLSRVSALPRDYGRLEVGTWRGLGNKLTIGGAMRYFGKXSIRATAERY 796
Db 805 YGVPECIDGGFPGGYLRTSIQPKYTANLVGGRLFDEKLELGSRLYHSRA--ENKDEKW 862
Qy 797 IDGTNGGNTSNFROLGKRISIKQTTTLARQPLIFDFYAAYPEKKNLIFRAEVKNLFDRIYI 856
Db 863 LMGV---LPNEYKGISNNPMRWNSVFT-----VDAYVSYQITPAISMELTGTNLNRYL 914
Qy 857 DPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNCKYGGTSGKSVLTNPFARQTELTIT 916
Db 915 DPL-----TRSMIP--APGRTFKLS 932
Qy 917 MSYKF 921
Db 933 LTSQF 937
```

Search completed: July 24, 2002, 08:58:34
Job time: 450 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 09:01:09 ; Search time 54.24 Seconds
(without alignments)
2273.886 Million cell updates/sec

Title: US-09-762-926-6
Perfect score: 4894
Sequence: 1 MRSEFLKPICFYLMGVTLX.....SVLNFARGRFLTIMSYK 921

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 410475 seqs, 133914855 residues

Total number of hits satisfying chosen parameters: 410475

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*
1: /cgn2.6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2.6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2.6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2.6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2.6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2.6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2.6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	275.5	5.6	917	5	US-09-584-501A-11
2	269	5.5	908	5	US-09-584-501A-12
3	231	4.7	934	5	US-09-540-209B-7679
4	204.5	4.2	564	5	US-09-545-199F-30
5	200.5	4.1	1083	5	US-09-540-209B-9521
6	200	4.1	925	5	US-09-540-209B-6249
7	194.5	4.0	967	5	US-09-545-199F-32
8	179.5	3.7	783	5	US-09-540-209B-9746
9	177	3.6	669	5	US-09-545-199F-105
10	176	3.6	707	5	US-09-540-209B-10058
11	173	3.5	1094	5	US-09-540-209B-8695
12	172.5	3.5	833	5	US-09-540-209B-7143
13	171.5	3.5	702	5	US-09-540-209B-8474
14	167	3.4	930	5	US-09-584-501A-2
15	161	3.3	760	5	US-09-540-209B-6609
16	159.5	3.3	1152	5	US-09-540-209B-10027
17	159	3.2	699	5	US-09-540-209B-6132
18	158.5	3.2	723	5	US-09-540-209B-6565
19	158.5	3.2	1119	5	US-09-540-209B-9958
20	157	3.2	1102	5	US-09-540-209B-7438
21	155	3.2	1008	5	US-09-540-209B-8074
22	154	3.1	947	5	US-09-540-209B-9753
23	154	3.1	1111	5	US-09-540-209B-9178
24	152.5	3.1	1102	5	US-09-540-209B-9537
25	152	3.1	722	5	US-09-540-209B-9749
26	151.5	3.1	1087	5	US-09-540-209B-7962

ALIGNMENTS

```
RESULT 1
US-09-584-501A-11
; Sequence 11, Application US/09584501A
; GENERAL INFORMATION:
; APPLICANT: Lo, Reggie Y. C.
; APPLICANT: Schryvers, Anthony B.
; TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS OF
; TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA AND VACCINES CONTAINING THE SAME
; FILE REFERENCE: A34763 021645.0106
; CURRENT APPLICATION NUMBER: US/09/584,501A
; CURRENT FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: 08/753,759
; PRIOR FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: CA 2,164,274
; PRIOR FILING DATE: 1995-12-01
; PRIOR APPLICATION NUMBER: 60/008,569
; PRIOR FILING DATE: 1995-12-01
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 917
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-09-584-501A-11
```

Query Match 5.6%; Score 275.5; DB 5; Length 917;
Best Local Similarity 21.1%; Pred. No. 7.4e-13;
Matches 221; Conservative 132; Mismatches 325; Indels 367; Gaps 57;

QY	5	FRKPTCFYLMGVTLHYSAEDAGRAGSEAQIOVLVDVHVAK--RVPDKKVFDTARA	62
Db	7	FRNLILSLMTAL---PAYAENV-QAG-QAQEKQLDITIQVAKKQKTRDNEVTGLGL	61
QY	63	VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGRVNTMVDGIT	118
Db	62	VKTADTLSKQVLDLRLTRYDPIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGCLA	117
QY	119	Q-TFYSTSDAG--RAGGSSQFGASVDSNFIAGLDVVVKGFSGSAGINSIAGSANLRTLG	175
Db	118	QIQSYTAQAALGGTRTAGSSGAINETIYENVKAVEISKGSNVEQSGCALAGSVAFTKT	177
QY	176	VDDVVOGNTYTGILLKLTGTNSTKGNMAAIGARKWLESASGVLYGHSRVSVAQNYR	235
Db	178	ADVVIGEGROWGT-----QSKTAYSGKNRGL-----TQSIALAGR	212
QY	236	VGGG-----GQHGNFGAEYLERRKORYFVQEGALKFN-----SDSGKWERDLQROQW	283

```
Db 213 IGAEAALLIIRTRHAGEIRAEHAAGR-----GVQSFNRLAPVDDGSKYAYFIVEEE- 263
QY 284 KYRKYNNNOEL-----OKYIEGHDK-----SWRENLAQVDITPIDPSSLKQOOS-----AG 331
Db 264 -----CKNGHEKCKANPKDVGDEKROTVSTRDYTCGNRFLA--DPLSVESRWLFRPG 317
QY 332 NLF--KLEYDGVNKNYTAQ---FRDLNWKIGSRKII-----NNRYQF-----N 369
Db 318 FRENKRHYIGGILERTQQTFTDRMTVPAFLTKAVEDANQKQAGSLRGNGNHYAGNHK 377
QY 370 YGSLNPNYTNLNTAAYNSGRQYKPGSKFTGWLGLLKDPETYNNAKLLDNTATRLPR 429
Db 378 YG-----GLTSGENNAAPGAAY-GTGVFYD----- 402
QY 430 ETELOTTLGFNPHNEXGKKNRPFPEELGFPD--GPQDNGCL----- 468
Db 403 ETHTKSRYLEVYVYNADKDTWADYARLSYDRQGLDNGHFFQTHCSADGSKYCRPSAD 462
QY 469 --YSYLGFRGDKGLLPQKSTIVQAGSYFNTFYFDAALKKDIYRLNYSNTVGY-REG 525
Db 463 KPFSY---YKSDRVIVGESHKLLQAAPKKSFTAKI-----RHNLNVN-LGYDRFG 509
QY 526 G-----EYTGYSDDDFKRAFGENSP----- 547
Db 510 SNLRHODY--YYOSAN---RAYSLKTPQNNKKTSPNGREKNPYWVSGRGNVVTROIC 564
QY 548 -----TYKKHCNRSC--GIYEPVKK-----YGKKRANNHSVS----- 578
Db 565 LFGNNYTDCTPRSINGKSYAARVDNRVLGRWADVAGLRYDYRSTHSDGGSVSTGTHR 624
QY 579 -ISADFGDYPMPFASYSRTH-----RMPNIQEMFYSQIGSGVHTALPKPERANTWQFG- 630
Db 625 TLSNAGILVKPADLWLDLYRTSTGFRLPSPFAEMYGWRSGDKIKAVKIDPEKSFNKEAGI 684
QY 631 -----FNTYKGLLKQDDTLGLKLVGYSRDN-----YTH----- 661
Db 685 VFKGDFGNLEASFENAYRDLIVR-----GYEAQIMDKREQVKGNGPAYLNAQSARI 735
QY 662 ---NVYGGK--W---WDLNGDIP--SWYSSTGLAYT-IQHRNFKDKVHKHGFELNLDYGR 711
Db 736 TGINILKIDWGVWD---KLPGWST--FAYNRVRVDIKKRADETDIQSHL-FD--- 786
QY 712 FFTNLSYAYOKSTOPTNFSDASESPNNASKEDQLKOGYGLSRVSALPRDYGRLEVGTWRL 771
Db 787 -----AIOPSRYV-----VGSYDQPEKGWGVNGMLT-----YSKAKEITELL 824
QY 772 GNKLITLGMARYGKSRATAEERYIDGTNGNTSNFRQLGKRSIKQETELARQPLIFDF 831
Db 825 GSRALLNG-----NSRNTKATARRTRPWYIVDV 852
QY 832 YAAYEPKKNLIFRAEVKNLFRDRIY 856
Db 853 SGYYTVKKHFTLRAGVYNLLNHRVY 877

RESULT 2
US-09-584-501A-12
; Sequence 12, Application US/09584501A
; GENERAL INFORMATION:
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Potter, Andrew A.
; TITLE OF INVENTION: PASTEURELLA BINDING PROTEINS OF
; PASTEURELLA HAEMOLYTICA AND VACCINES CONTAINING THE SAME
; FILE REFERENCE: A34763 021645.0106
; CURRENT APPLICATION NUMBER: US/09/584, 501A
; CURRENT FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: 08/753,759
; PRIOR FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: CA 2,164,274
; PRIOR FILING DATE: 1995-12-01
; PRIOR APPLICATION NUMBER: 60/008,569
; PRIOR FILING DATE: 1995-12-01
```

```
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 908
; TYPE: PRX
; ORGANISM: Neisseria meningitidis
US-09-584-501A-12

Query Match 5.5%; Score 269; DB 5; Length 908;
Best Local Similarity 21.2%; Pred. No. 2.3e-12;
Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;

QY 5 FRLKPICFYLMGVTLHYSYAEDAGRAGSEAOIQVLEDVHVRAK--RVPKDKKVFTDARA 62
Db 7 FRLNLCLSLMTALPV--YAENV--QAEOAEQKOLDTIQVRAKKQKTRREDNEVTGLGKL 61
QY 63 VSTRQDIFKSS-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGSGFRVNTVMDGIT 118
Db 62 VKSSDTLSKEQVLNIRDLTRYDPGIAVVEQGRGASSG--YSIRG-MDKNRVSLTVDGVS 117
QY 119 Q-TFYSTSTDAG--RAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSILAGSANLRTL 175
Db 118 QIOSYTAQAALGTRTAGSSGAINIEYENKAVEISKGSNSSEYNGALAGSVAFTQKT 177
QY 176 VDDVVOGNNYTGILLK-GLTGTNSTKGNMAAIGARKWLESASGVGLYCHSRRSVAQNY 234
Db 178 AADIIGEGQWGIQSTAYSGKDHALTQSILAGR---SGGAELLIIYTKRGREIHAAH 233
QY 235 RVGGGQHQHGNFGAEYLERKQ-----RYVOEGAL-----KFNSD-SGKWER-- 276
Db 234 KDAGKG--VQSFNRLVDEDKKEGSGYRYFIVEECHNGYAAACKNKLKEDASVKDERKT 291
QY 277 -----DLQROQWKYP-----YKNNNOELQKYIEGHDKSWRENLAPO 314
Db 292 VSTQDVTGSNRLANPLEYGSQSWLFRPGWHLNDRHYVGAVLERTQOTEDT--RDMTVPA 349
QY 315 YDITPID--PSSLK-----QOSAGNLFKLEY-DGVFNRYTAQFRDLNFK 355
Db 350 Y-FTSEDYVPSLKGKLYSGDNKAERLFVQGGSTLQIGIGYGVF-----YDERHTK 402
QY 356 IGSRKIIINRYQPNYGLSLNPTNLNLTAAYNSGRQYKPGSKFTGWLGLLKDEFTNNAK 415
Db 403 -----NR-----YGVY-VYHNADKDTWADYARLSYDRQ----- 431
QY 416 ILDLNNTATFRLPRETELQTLGTFNHFNEYGKRNPEELGLFFDGDQDNGLYSYLGRF 475
Db 432 -IDLN-----RLQOT---HCSHDGSDKNCRP-----DG-----NKPYSF--Y 463
QY 476 KDGKGLLPQKSTIVQAGSYFNT-----FYFDAAL----- 506
Db 464 KSDRMIEESRNLFOAVFKAFDTAKIRHNLINLGYDRFKSQLSHSDYVQLQNAVQYDL 523
QY 507 -----KKDIYRLNYSNTVVG---YRFGGE-YT-----GYGSDDEPK 539
Db 524 ITPKPPFPNGSKDNPYRVSIGKTTVTSPICRFGNNTYDCTPRNIGNGNGYAAVQDNV 583
QY 540 RAFGENSPYKHKCNESCGIYEPVLKKYKKRANNHSVS-----ISADFGDYPMPFAS 593
Db 584 R-----LGRWADVAGIRYDYRSTHSEDSVSTGTHRNLSWAGVYVLPKFTWM 631
QY 594 SRTH-----RMPNIOEMFYSQIGSGVHTALPKPERANTWQFG-----FNT 633
Db 632 DLTyrASTGFRLPSPFAEMYCWARGESLKTLDLKPEKSFNREAGIVFKDGFNLEASYFNN 691
QY 634 YKGLLKQDDTLGLKLVGYRSRDN-----YIH-----NVYGGK--WDLNG 672
Db 692 AYRDLI-----AFGYETRTQNGQTSASGDPGYRNAQNAIAGINILGKIDWGVWG 742
QY 673 DIPSWVSSSTGLAY-----TIQHRNFKDKVHKHGFELNLDY--GREFTNL 716
Db 743 GUPDGLYST-LAYNRKIVKADIRADRTFVTSYLFDAVOPSRVYVGLGLGHDHDPGIGINT 801
```

539	Db	KLKVGDIIVYRDFDGVMSGVFAQ-----GEYNRDKLS	571
577	Qy	VSIADFGDYFMPFASYSRTHRMPIQEMFYSGIDSGVHTALKPERANTWQF--GFNTYK	635
572	Db	AFISGSYSN-----TGYWRYDR-----FYDK-----AHA-----KSKTVNEIGNNA--	608
636	Qy	KGLLKQDDTTLGLKL---VGRSRIIDNVIHNVYKQWDLNGDIPSWWSSTGLATYIQRNF	692
609	Db	KGGLNYNLTEHNHNFAVIGYISR-----APFFSGGAPLNSTVSNATN	650
693	Qy	KDKVHKHGFEELELNDYDGRFFTNLSY--AYO-----KSTOPTNFSDASES--PNNASK	741
651	Db	PDVAVNEKFSEEGYGRSSEFLVININAYHTRMWDKTTTTRSQDITINYEGSLSEPDAK	710
742	Qy	EDQLKQGYGLSRVSALPR-----DY-----GRLEYGT--RWLGKNKITLGGAMRYFG	785
711	Db	LVSTKSVINMGVNALLHQGVLEDFVAKPFQWLDLSGMFSGIGNRWDSN---ASGSFTVEG	767
786	Qy	KSIIRATAERYIDTNGGNTS-----	806
768	Db	QFVNSAS-----IKGSDGKDVTLVNVAAANGLEPGTMKLNKDKVKGSGAQTAAALGATEK	823
807	Qy	-----NFRGLGKRISKQETTLARQPL--IFFDYAAVE---PK	838
824	Db	IDKALRGIDWNLYARNYADWSLNSNDLVNSEKDFSTPNRIPTASTFDLNASYKKNFGK	863
839	Qy	KNLIIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSF	874
884	Db	INAVISGNVNI.PQOTVISIDATDGSNNHDKWTAYNVF	919

```

RESULT      4
US-09-545-199F-30
; Sequence 30, Application US/09545199F
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David
; APPLICANT: Fuller E., Troy
; APPLICANT: Kennedy J., Michael
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/6227.NCP
; CURRENT APPLICATION NUMBER: US/09/545, 199F
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/153, 453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128, 689
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-09-545-199F-30

Query Match          4.2%; Score 204.5; DB 5; Length 564;
Best Local Similarity 20.0%; Pred. No. 1.3e-07;
Matches 131; Conservative 87; Mismatches 271; Indels 165; Gaps

Qy    41 EDVHVKARVPKDKKVFTDARAVSTRQDIKFSSSENLDNIVRSIPCAFTQQDKSGGIVSLN 100
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    44 EDNKLPGRSVLKQRNIQ-----QQADNAADLNILPEVNWAGGFPGGQTIN 91

Qy    101 IRGDSGGFRVNTVWDGITQTFTYSTDAGRAGGSQFGASVDSNFIAGLDVVKGFSFGSA 160
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    92 INGMGDADRVQLDGATKSPEK-----YQGGSIFTEPELLRKVTVDKGNYSPOY 141

Qy    161 GINSLAGSANLRLTGLGVDDVVGNNNTYGILLKLGTNS----TKGNAMAIGARKWLSE 215
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    142 CNGGFAGTVHKETKDATDFLEKENOKIGGLFK--YGNNSNNOKTYSTALVLQNEKNIDL 199

Qy    216 GASVGYLVGHSHRRSVAQNRY-----VGGGGQHIG-----NFGAEYLERRKQRYFYVQ 261
      ::||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

Db 200 -----LLFGSVRN--ASNYTRPDKSKILFSKNNQKSLIKVNNQITPEHLLTLSSVYGIH 252
Qy 262 EGALKFNSDSKWERDLQROQWKYKPKYKNYNQELQYIEGHDKSWRENLAPOYDITPID 321
Db 253 KG-----WEPWAARQVMSRP-----TETBIKH--GIDVAKRKL-- 287
Qy 322 PSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKLIINRYQFNGLSLNPTNLN 381
Db 288 -----YRD-----QKDESYSLKRYLPENKNWINLS 313
Qy 382 LTAAYNSGRQYKPKSGKFTG--WGLL--KOFETYNNAKILDLNNTATFRLPRETELOTTL 437
Db 314 VOLSYSKTEQNDTRHEKVTFSFLGTGKNSWITYSDL--TFDISNTSLNIGR-AEHELLE 371
Qy 438 GFNFHNEYGNKRRPEELGLFFDDPDODNGLYSYLGRFKGDKGLLPQKSTIVQAGSOYF 497
Db 372 GLOWLKNK--RNTLMYHKG----GVKKADYNYGFQY-----YMPSGRQYT 412
Qy 498 NTFYFDAALKDIYRLNYSNTVGYRFGEGYTYGYSDDDEFKRAFGEN--SPTYKKHCHNRS 556
Db 413 QAEYLOQDIKWONFL-----TGGIRY-----DHINNIGOKNLAPRYN--DIS 453
Qy 557 CGIYEPVLKYGKGRANNHS--VSIADFGDYFMPFASYSRTHRMPNIQEMY----FSQIG 611
Db 454 AG-----HDYSQKNYNGWSYILGLKYDVNHYLSLFTNFSKTRAPVIDEQYETQYSQAS 507
Qy 612 DSGVHTALKPRANTWOFENTYKKGKLLKODDTLGLKLVGRSRIIDNHYHNVYG 665
Db 508 VSATSLNEREMINQTVGGIITLNLHLEQENDAFQFTTYFYNRCKNEIFKTRG 561
RESULT 5
US-09-540-209B-9521
; Sequence 9521, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRA
; FILE REFERENCE: 2709,1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 9521
; LENGTH: 1083
; TYPE: PRP
; ORGANISM: B.fragilis
US-09-540-209B-9521

Query Match 4.1%; Score 200.5; DB 5; Length 1083;
Best Local Similarity 20.2%; Pred. No. 6.8e-07;
Matches 197; Conservative 129; Mismatches 369; Indels 279; Gaps 57;
Qy 4 SFLRKPICFYLGMVTLHYHVAEDAGRAGSEAQIOVL--EDVHVAKRV---PKDKKV- 56
Db 86 SLKVAPNSQFLVS-----YIGKQOTIKVSESTYNIKVEDAEVLDEVVVGVSQKKNV 141
Qy 57 -----FTDARAVSTRQDIFKSENLENIVRSIPGAFTQODKS-----SGIVSLNIRGDSG 106
Db 142 VTGAVGVMSAEALR--PVANASALQGVV---PGLNLTVNNGGALDGLTNMIRGACT 197
Qy 107 FGRVN-----TWDGITOFTYSTSD-----AGRAGGSQFGASVDSNFIAGLDVVKG 154
Db 198 IGDGSSSPLVLIDIGEGDLNTPNDIESVSVLKDAASAIYGARASFGVI--LVTTKS 255
Qy 155 SFGSAGINSLAGSANLR--TLGVDDVVQGNNTYCLLLKGLTGINTKGN-----AMAA 206
Db 256 GKSKNTNV--SYSGSARFSAIDGVPDIM---DSY--TFAQYFNRAANKGGGDFAPAYMER 310
Qy 207 IGA-----RKWLESASVG---VLYGHSRRSVAQNYRVGGGQGHGN 245
Db 311 IKAYQEGTLKATTVYDNGAGIWKWANANGDTDWFEFYDHWASQEHNLNSINGTD----- 366

Qy 246 FGAELYERRKQRYFV-----QEGALKFNSD-----SGK-----WERDLQROQWKY 285
Db 367 -----KTQYLISGSFLDQKGLMRHGDKFQRYTLNGKIITAVTDWFKVYISTKWR 417
Qy 286 KPKYKNYNQELQYIEG---HDKSWRENLAPOYDIT--PIDPSSLKQOSAGNLFKLEYDG 340
Db 418 EDF-----ERPSTLTGNFFHNLARKVPVHPAYDPNGFPMDGEVEQEMNGGKQNSQDF 471
Qy 341 VFNKYTAQFRDL--NTKIGSRKIINRYQFNGLSLNPTNLNLTAAVNSGRQYKPKSGK 399
Db 472 YTNOLQVLPFIPKKNKINLDSVRTTQYQHWEVLVY-----AYNAVADPY-----Y 519
Qy 400 TGWGLLKDEFTYNNAKILDLNNTATFRLPRETELOTTLGFNFYHNEYGKRRFPEELGLFF 459
Db 520 TVMDMGVGYAAGSSRVNE-----YSWKENYTTTNIYSDFYKQF- 558
Qy 460 DGPDDQNGLY--SYLGRPKGD---KGLLPQKSTIVQAGSOYFNTF-----YDAA 505
Db 559 -----DNGHYFKVMAGFNABLYKTRNITAEKNTLITP--GVPTINTATDDPQAYGGVADNS 612
Qy 506 LKDDIYRLNYSNTVGYRFGEGYTYGYSDDDEF--KRAFGENSPTYKKHCHNRSQGIYEPV 563
Db 613 VAGFFARVNNYSKD---RYMFEANGRYDGSRRFVGKERWG--FPFSFSAGWNIA---REPF 665
Qy 564 LKYGKGRANNHNSVISISADFG-----DYFMPFASYSRTHRMPNIQEMYFSQIGDSGV 615
Db 666 MESFAEK--INMGLKLRASWGQLGNTNTDAWYF-----YQTMP--VGSNYGWLNVGERP 718
Qy 616 HTALKP-----BRANTWQFG-----FNTYKKGKLLKODDTLGLKLVGRSRIIDNHYH 661
Db 719 NYATNFGIVSSKKTWETVETWVGLWDSFFNNRLSG-----SFDYFVR 761
Qy 662 NVYKGWDLNGDIPSWVSSYGLAYTIOHRNFKDKVHKHGHFELELNY-----DYGRFFT 714
Db 762 YTY----DMIGPAPELSSLLGT--SVPKINNDS--MKSYGFELEVNWRDRDICEVSYGAKFV 814
Qy 715 NLSYAVQKSTQPTNFSASPPNASKEDOLKOGYGLSRV-----SALPRDYGRLE 765
Db 815 --LSDQOQKILR---YPNDSYDVGSYKGEHLNDIWLTTTIGIAKSQBEEMDAHLAKVDSS 870
Qy 766 VGTFRW--LG-----NKLTLGAMRVFGKSIKIRATAERY---IDGT--NGG 803
Db 871 VGTNWVGDIIMYADLDGDGKISNGTNKLGTDGYRIIGNS---TPRFKYGITLDAWKGF 927
Qy 804 NTSNFRQ--LGKRSI 816
Db 928 DFSIFMQGIGKRD 941
RESULT 6
US-09-540-209B-6249
; Sequence 6249, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES
; FILE REFERENCE: 2709,1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6249
; LENGTH: 925
; TYPE: PRP
; ORGANISM: B.fragilis
US-09-540-209B-6249

Query Match 4.1%; Score 200; DB 5; Length 925;
Best Local Similarity 19.8%; Pred. No. 5.9e-07;
Matches 154; Conservative 107; Mismatches 257; Indels 260; Gaps 40;
Qy 34 EAQIOVLVDVHVAKRVKPKDKKVFETDARAVSTRQ-----DIFKSENLDNIVRS 82

```
Db 236 EETHMLDEVITISGRIONVKSTOLGAETLRPTQLKNIPMALGEVDILK-----MVQA 288
Qy 83 IPGFTOODKSSGIVSINIRGDSGFGKRVNTWVGITOTFYSTS-----T 126
Db 289 LPGVKTYGEASSG---ENVRGGATDQNLILLNDG---TIYNPNHLFGFFFAAFNSDMVKEA 342
Qy 127 DAGRAGSSGFGASVDSNF--TAGLDVVVKGSFGSAGINSLAGSANLRTLGVDG---VVQ 181
Db 343 EYKSSTPAQGGRISSILDTKEANKERTGSGTGLVTSKLNLEIPIKORTSVLLS 402
Qy 182 GNNYTG-LLKGLTGNTSK-GNA---MAAIGARKWLES-GASVGLYGHSHRRSVAQNY 234
Db 403 GRTTYSWIMKOLPEKSEYKNGTAGFYDLAAIVAHKENDKHSNLNVIYGYSHDRFAFNSNE 462
Qy 235 RVGGGGHICNFGAELERRKQRYFVQEGALKFNDSGKWERDLQROQWKYKPKYNNQ 294
Db 463 KYG-----YNNLNA-----SARWR-----AVNEK 482
Qy 295 ELQKYEIGHDKSWRENLAPOYDITPIDPSLKKOOSAGNLFKLEVDGVFNKYTAQFRDLNT 354
Db 483 LIGYFSAGY-----HYD-----NRETVNASTAKLSFD--INQTF-----V 519
Qy 355 KIGSRKIINRYOFNYGLSNPYTNLNLTAAYNSGROKYPKSGKFTGWLKDKDFETYNNA 414
Db 520 KADFTNILADKHTLNFGEK-----SMLYHINSGTVE-PEGS-----ESFVK 560
Qy 415 KILDNNT--ATFPLPRETELOTTGLGNFYHNEYGKNRFPPEELGFLFDPDQDNGLYSYL 472
Db 561 DVLOKDALETAFLVGDWEITPKLSVN-----AGIRY-----SLFSAL 599
Qy 473 G---RFGKDGGLLPKOKSTI--VQPAGSOYENTFYDAAALKKDIYRLNYSNTVGYRFGG 526
Db 600 GPRSYOYAGMLPHESTIITDTITAGAGKFMKTYH-----GP 636
Qy 527 EY--TGYGSDDEPKAFGENS-----PTYK-----550
Db 637 EFRLSARYAFTDNFVSKAGNSMRQYIHKLSNTVIMSPDTWKLSDVNIKPRQGWQAAG 696
Qy 551 KHCNRSGIPEVLKYKGRKRNHNSVSIADFGDYPMPFASRTRHRMNIQWYFSQI 610
Db 697 LYLNSPGIWEYSVEGYK-----MSDYLDY-RGGAKLLMNHIIET--DVINTQG 744
Qy 611 GDSVHTALAPE--RANTWQFGFNTYKGLLKODDITGLKLVGRSRIIDNYIHNVYKWW 668
Db 745 HAYGVELOVKQVOKLNGWM--SVTYSRTFLRQND-----KRIEKPVNN--GDWY 790
Qy 669 DLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFELNLYDYGFRFT--NLSVATQKSTQ 725
Db 791 PTEYDKPHDKFVG-----NVK-FTHRYSMSINVDYSTGRPTTIPAGQYVDESTQ 839

RESULT 7
US-09-545-199F-32
; Sequence 32, Application US/09545199F
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David
; APPLICANT: Fuller E., Troy
; APPLICANT: Kennedy J., Michael
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/6227.NCP
; CURRENT APPLICATION NUMBER: US/09/545,199F
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 967
; TYPE: PRF
; ORGANISM: Pasteurella multocida
US-09-545-199F-32
```

```
Query Match 4.0%; Score 194.5; DB 5; Length 967;
Best Local Similarity 19.0%; Pred. No. 1.7e-06;
Matches 207; Conservative 131; Mismatches 339; Indels 413; Gaps 51;
```

```
Qy 39 VLEDVHVHAKRVPKDKVFTDARAVSTRODIFKSSSE-NLDNIVRSIPCAFTQODKSSGIV 97
Db 22 ILADSHOEATEL-----DTITVSSQDDEMTKEKIGETVKTASOLKQKQVQDSDRL 73
Qy 98 SLNIRGDSGFGRVNTWVGITOTFYSTSDAGRAGSSGQFCASVDSNFIA-----GLDVV- 152
Db 74 --VRYETGV-----TVVEAGRFGSGYAIRGVNDENRVAITVDGLHQA 114
Qy 153 -----KGSFSGSAGINSLAGSANLRTLGVDVVOGNNTYGLLLKGLTGNTSKG----- 201
Db 115 TLSQSGKELFEGYGNFNTRNSVEIETLVAKIAGADSVK-----GSGSLGAVLF 168
Qy 202 --NAMAAGARKW--LESASGVLYGHSHRRSVAQNYRVGGGQHGIGHGAEYLERK 255
Db 169 ETKDARDFLTEKDMHIGYKAGYSTADNOGLNAVTLAGRYQM-----FDALIMHSKR 219
Qy 256 QRYFVQEGALKFNDSGKWERDLQ--ROQWKYKPKYKNYNNQELQY-----299
Db 220 HGHELE---NYDYKNG---RDIQKEREKADPTIITKESTLVKFSFSPTEHHRFTVASD 272
Qy 300 -----IEGHDKSW-----RENLA---POYDITPI-----320
Db 273 TYLQHSRGHDLSYNLVATTHIQLDEKESRHANDLTKRKNVSFTYENTVTPFMDTKLSY 332
Qy 321 -----DPSSLKKQOS-----AGNLFKLEYDGVFNKYTAQF 349
Db 333 SQORITTRATEDYCDNELCDSYKNPLGLQFDKGQILDPAKNKIKOGSGL-----STQI 388
Qy 350 RDNLTKIGSRKIIINRYQNFYGLSL-----NPTYNLMLTAAYNSGRQ---391
Db 389 VDENGKFPPTTGTGNNAAFNNLRLRTGFWLDCVFDCKNPFVTVYINSNGTYOAREVLL 448
Qy 392 -----KYPKSGKFTGWL-----LKDPETYNNAKILDLNNTATFRL 427
Db 449 SEETVQDKLYKTAKEEG-GLPNYLILPNKGYLPYDYKERDLNNTKQINLDTKTF--505
Qy 428 PRETELTGLGFNYFHN-EYG--KNRPEEL--GLFEDG-----PDODNGLSYLG--R 474
Db 506 -----LTFNLNLSYGVGYVSRIEKEMINKAGYEGRNPTWADRILGQSSYCYGNA 556
Qy 475 FKGDKG-----LILPQKSTIVQAGSOYFNTFYDAAKKDIYRLNYSNTVGYR-----523
Db 557 LKCPKHEPLTSFLIPVEATT-----QSLYFANILKVH---NMISIDLGYRYDHIK 603
Qy 524 FGGEYT-----GYI-----GSDDEFKRA-FGENSPYKHKHNRSCGIYE 561
Db 604 YNPEYTPGVTPKIPDDMVKGLFIPMPKEPOLKDPDYNAKFGAYKKWKEYLPKNA--EE 661
Qy 562 PVLKYYKGRKRNHNSVSIADFG--DYFMPFASYSRTRHRMNIQWYFS-QIGDSGV--H 616
Db 662 NIAYIAQDKTKFKHSYSLGATFDFPLNFRVQVYKYSKGFRAPTSDELYFTFKHPDFTILPN 721
Qy 617 TALKPERANTWQFGFNTYKGLLKODDITGLKLVGRSRIIDNYIHNVYKWWDLNGDIPS 676
Db 722 PVLKPEEAKNOETALTVDHNNGFVSTSVFQTK---YRHFID-----LA 761
Qy 677 WVSSTGLAYTI---QHRNFK-----DKVHKHGFELNLYDYGFRF-----TNLSY 718
Db 762 YLGSRLNSVSGQAQARDQVYQNVVNDNAKVGLKLEINARNLNGLYFHWHLGDPNTSYKF 821
Qy 719 AYQKSTOPTNFSASESPNNA-----SKEDQLKQGYGLSRVS-----A 756
Db 822 TYORGLRD-----GDRPMNAIQKASVFGLYGDHKNFKTGADLYITRVSEKKAADPYNM 875
Qy 757 LPRDYGRLEVGTRWLGKNTLGGAMRYFGKSIRATAEERVIDGNTGNTSNFRQLGKRSI 816
Db 876 FYKEQGYKDSAVRWRSDDYTLVDVAVGI-----903
```

QY 817 KOTETLARQPLIEDFYAAEPKKNLIFRAEYKKNLFDRRY-----IDPLDAGN----- 863
Db 904 -----RPKNLTLOFGVYNLTDRKYLTWESARSIKPFGTSLINQK 944
QY 864 -DAATQRYYS 872
Db 945 TGAGINRFYS 954
RESULT 8
US-09-540-209B-9746
; Sequence 9746, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES PRA
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540, 209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 9746
; LENGTH: 783
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-9746

Query Match 3.7%; Score 179.5; DB 5; Length 783;
Best Local Similarity 20.8%; Pred. No. 1.9e-05;
Matches 139; Conservative 84; Mismatches 251; Indels 193; Gaps 30;
QY 34 EAQIQVLEDVHVAKRVPKDKVFTDARAVSTDRIFPKSSNLDNIVRSIPGAFQTQODKS 93
Db 107 EPTQSLGSEVVVYAKSEARQLREQAPMVSIMQOLOGYNSVDVLSKTVGVTIRTTGG 166
QY 94 SGIVS-UNIRGDSGFRVNTWVGITOTFYSTSTDRAGRAGSSQFGASVDSNFIAGLDVV 152
Db 167 VGSRRSVVRGLEG-KRIGFFIDGSPMNDNSDFIDN-----DIPVDMIDRIEY 215
QY 153 KGSFSGSAGINSLAGSANL-----RTLGVDVVOGNNYTGILLKGLTCTNSTKGNAMA 205
Db 216 KGVVPAREGSSVGGAVNIVREPPKYLDASYSIESFNTHKLSL-----VTKRNI-- 266
QY 206 AIGARKWLESASGVLYGHSRRSVAQNYRVGGGQHIGNFGAEYLERRRKQRYFVQ--E 262
Db 267 --ATKGLFEGG--GGFYTS--DNNYKME-----SPFEEGLIIRKNDKFKKLAVA 311
QY 263 GALKFNDSKWERDLQROQWKYP-----YK-----NYNNQELQKY----- 299
Db 312 GSLK---ARKWFDLA---EFEPVFIHTFKETQIEYNIERAKHTYSDAFIFANKLEKE 363
QY 300 ---IEGHDKSWRENLAPOYDI-TPIDPSSLKQOSAGNLFKL--EYDGVFNKYTAQFDLN 353
Db 364 NFLTEGLD--WESMLAYATVFHMVDTAAYRNWDGTYPAVSEYGEIGKWNASARN-- 419
QY 354 TKIGSRKIINRNYQFGLSNLYTNLNTAAVNSGRQKYPKSKFTG----- 402
Db 420 ----EKHTIHKHLNVINNHSINLSLFSASGHPKDDLKNKVGVYKYNFSTMASW 475
QY 403 --GLLKDFETYNNAKILDLN-----NT-----ATFRLPRETELOTTLGFN 441
Db 476 IAGLGYDFERTDNDIFLSNLNVKYYMYGMNTHMSSINSSEAEKVDMLKRDGFSNALRYR 535
QY 442 FHNEYGK-----NRPFEELGLFEDGPDQDNGLYSYLGRFKGKGLLPKQSTIVQAGSQ 495
Db 536 TPDFMGKLSVGVYDVRLEPAESSELGDC-----YTVAPSGN--LLPERNTSV----- 578
QY 496 YFNTFYDAALKKDIYRLNYSNTVGVYFGGVTGYGSGDEDFKRAFGENSPYKKHCNR 555
Db 579 --NLGFLLDRTKXASNLQVEVNT-----FYGYLENMIRFTG----- 613
QY 556 SCGIYEPVLKKGKRRANNHSVSISADFGYFMPFASYS-----RTH----- 597

Db 614 --CYLOSQYQNECKMRTLGVEVEVKADLTHWLYGYCNMTYQDLRDVRKKEPNTHTNPTK 671
QY 598 --RMPNI 602
Db 672 GSRMPNI 678
RESULT 9
US-09-545-199F-105
; Sequence 105, Application US/09545199F
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David
; APPLICANT: Fuller E., Troy
; APPLICANT: Kennedy J., Michael
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/6227.NCP
; CURRENT APPLICATION NUMBER: US/09/545,199F
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 105
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-09-545-199F-105
Query Match 3.6%; Score 177; DB 5; Length 669;
Best Local Similarity 20.3%; Pred. No. 2.3e-05;
Matches 170; Conservative 95; Mismatches 296; Indels 276; Gaps 40;
QY 109 RVMTWVDGITQTFYSTDA-----GRAGSSQFGASVDSNFIAGLDVYVKGFSFGSAGINS 164
Db 8 RVAVIVDGIPIQAEISTISARSYSTERHNGNIN--NIEVENVSSLKVKQGAASVMTYGSA 64
QY 165 LAGSANLRTLGVDVVOGNNYTGILLKGLTGTNSTKGNAMAIG-----AR 210
Db 65 LGSTVETFTKDIEDFVPEGRHGLFSK--FGYTSKNREYRQVIGVGKGGEHFGFVOLT 122
QY 211 KW----LESASGVLYGHSRRSVAQNY-----RVGGGQHIGNFGAEYLERRRKQRYF 259
Db 123 RWGHETINNGKGTDLGEHRGKPNLNYITTSWLTQVGYDINNTHRF-TLFLDREKKL 181
QY 260 VQEGALKFNDSKWERDLQROQWKYPKYNYNNOELQKYEIEGHDKSWREN-----LAPQY 315
Db 182 TEEKTLGL-SDAVRFAND-----QTPYLRYGIE--YRY--NGLSWLETVKLFLAKQ- 227
QY 316 DITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFDRLNTKIGSRKIINRNYQFNYGLSLN 375
Db 228 -----KIEORSALQEF-----DIN-----NRN----- 244
QY 376 PYTNLNLTAAY-NSGRQKYPKSKFTG---WG-----LLKDFETYNNAKILDLNNTATFR 426
Db 245 ---KLDSTMSFVYLOQNTARGESTSPYWGSPSRHLSAKFE-FRDKFLNNKHHFTFR 300
QY 427 LPRETELOTTLGPNYFNHNEYKGNRFPPELGLFFDQDNGLYSYLGRFKGDKGLLPQKS 486
Db 301 -PWQ-----INRFRQO--GRNNYTE-----LGRWDHY----- 317
QY 487 TIYQVAGSQYFNTFYFDAALKKDIYRLNYSNTVGVYFGGVTGYGSGDEDFKRAFGENS 546
Db 318 --VFPVKRSFEFSLSLDDDKIGELLHLG-----LGRWDHY----- 351
QY 547 PTYKHCNRCSGIYEPVLKKGKRRANNHSVSISADFGYFMPFASY--SRTHRMPIIOE 604
Db 352 -NYKPLNLSQHNINRQRLPY-PKTSKFSYQLSLEYQLHPSHOIAYRLSTGERVPRVED 409
QY 605 MYFSQIGDSG-----VHTALKPERANTWQFGFNTYKGLLKKODDTGLKLVGYRS 654

```
Db 410 LYFEDRCKSSQFLPNPDLOPEALNHEISYRFQNOYAHESVGL-----FRT 456
Qy 655 RIDNYIHNVYKWWDLNCDIPSWVSSTGLAYTIOHRNFK-DKVHKGHFELELANYDGR-- 711
Db 457 RYHNF-----IQRREMTCDKI-----PYEYNRTYCYCT 484
Qy 712 -----FETNLSYAVOKSTOPTNFSDAESPNNASKEDQLKQGYGLSRVLSALPRDYGRLEV 766
Db 485 HNTYVMEVNEPEAVIKGVEVSGALNGSAGLSDGLTFLRLKGSYSKGO-----NHGDPLKS 540
Qy 767 GTRWL-----GNKLTGLGAMRY-FGKSTRATAERY-IDGTNGGNTSNFRQLGKRS 815
Db 541 IQPTVVVTGIDYETEGHSVLSG--RYSAAKAKADAETEYTH-----DKKV 585
Qy 816 IKOTETLAROPLIFDYAAAYEPKKNLFRAEVKNLEDRYIDPLDAGNDAATQRIYS 872
Db 586 VKQPHLSPSYFVVDFTGQVNLKSNVILNMGVENLFNRDYM-TWDSAYNLTRGYTS 641

RESULT 10
US-09-540-209B-10058
; Sequence 10058, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 10058
; LENGTH: 707
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-10058
```

```
Query Match 3.6%; Score 176; DB 5; Length 707;
Best Local Similarity 20.7%; Pred. No. 3e-05;
Matches 174; Conservative 94; Mismatches 333; Indels 238; Gaps 47;

Qy 36 QIOVLEDVHVYKARVPKDKKVFVDARAVSTRODIFKSSSENLDNIVRSIPGAFQQDKSSG 95
Db 1 EVQI-----VSTRATSTPVAFTNVSKEELKQNF--GODIPFLLSMTPSALATSDAGAG 53
Qy 96 I--VSLNIRGDSGFRVNTWVDGTTQ-----FYSTSTDAGRAGGSGSOFGASVDSNEI 146
Db 54 IGYTTLVRGTG-D-TRINITANGIPMDAESHTLFWNMPD-----PASSVKD--- 100
Qy 147 AGLDVKVSGSGSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGTNSTKGNAMAA 206
Db 101 --IQVORGACTSTNGAGAFGASVNMQTEGI-----SMQPYAEINASVDSFNAHKETVKFG 153
Qy 207 IGARK--WLESG--ASGVLYGHRSRV-AQNTRVGGGGHIGNFGAEYLERKQRFTVQ 261
Db 154 TGLLKDHAFDARLSTGTGIDRASVDLYSEYAQGG-----YFAD 195
Qy 262 EGALKNSDSGK-----WERDLQROQWYKYP-----YKNYNQOELQ- 297
Db 196 NTSVKFTTFGGEKTYHAWYATKEEMKKYGPREFNSCCMTDHHGHRFYKQDQDNVLMQ 255
Qy 298 --KYIEGH--DKSWRENLAPOYDITPID-----PSSLKQOSAGNLFKLEYDGVFNKY 345
Db 256 NYQLLNHTESAANLNAALHY--TKGDYGYQYKEDRSIKEY---RLHPFMYDG---K 306
Qy 346 TAQPRDLNTKIGRKIIINRYQFNGLSLNPYTNLNLTAAYNSGRQRYPKGSFTGW--- 402
Db 307 EVEKSDL---IROKKMDN---HFGGGVFSVNYPHQNNDASLVNALNY-----DGWHFG 354
Qy 403 -----GLLKDEFTY--NNAKILD--LNNTATRLPRETELOTTLGFNYPHNE-YGK 448
Db 355 RVITWKNYIGELLPDHEYRNKAKKTDGNLYLKANYNLVAGLNAYADLQRYINYKTHG- 413
```

```
RESULT 11
US-09-540-209B-8695
; Sequence 8695, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 8695
; LENGTH: 1094
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-8695
```

```
Query Match 3.5%; Score 173; DB 5; Length 1094;
Best Local Similarity 20.4%; Pred. No. 9.7e-05;
Matches 174; Conservative 114; Mismatches 288; Indels 276; Gaps 48;

Qy 52 KDKKVFVDARAVSTRQDIFKSSSENLDNIVRSIPGAFQQDKSSGIVSLNI--RGDSGFR 109
Db 100 KDKYIM-----IIPQSKVEVESKKLSGIVK-----DDKGPLIGVNVSFKG-SPTGT 145
Qy 110 VNTWVDG-----ITQTFYSTSTDAGRAGGSSQF-----GASVDSNFIAGLDV 151
Db 146 V-TGLDGRFSILAAGNIIEFSYGVYTTQYITVGDASSLFVLEEDAKALDEVVVVTALGI 204
Qy 152 VKGSFSGSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLK--GLTGTNSTKGNAMAAIGA 209
Db 205 KRAEKALSYVQVKSDA-----INDVKDANFVNLTKGVAGVSNRSSSGIG---GA 254
Qy 210 RKWLESAS--VG---VLY-----GHSRRSVAQNTRVGGGGHIGNFGAEYLERK- 255
Db 255 TRVYMRGAKSIVGNNNLYVVDGMPIGNPSKGEINNDSTPGGEGISDFNPEDIESLSI 314
Qy 256 -----QRYFVQEGALAFNSDSGKWERDLQROQWYKYPKYNNOELQKI----- 300
Db 315 LTGPAAALYGVSSAANGVILINTKKG-----QEGKLLISINNTFETPYVMPFQN 366
Qy 301 -----EGHDKSWRENLAPOYDITPIDPSSLKQ-----QSAGNLFKLEYDGVFNKYTAQFRD 351
Db 367 RYGNNAKGSYKSWGEMLO-----QSFTRPKDFKFTGANIMNAANFSVGNKNNQTFVS 418
Qy 352 LNTKIGSRKII--NRNYQFNVGL-----SLNPYTNLNLTAAY-----NSGR----- 390
```

Db	419	VAT-TNSTGIIIPNNEYRYNFTLRNTASMLNDKHLDLGASVYLOQDQWLSAGRYFNPL	477
Qy	391	--QKYPKSGKFTGWGLLKDFETYNNAKILDLUNNTATFRLPRETELQTTLGFNYFHNEY-	446
Db	478	VPLYLEPRGEDFEA---VKYERY-----DTNR-----KPIQEWISYDQGLN-LENPYW	523
Qy	447	-----CKNRFPEELGFFDGPDDQDNGLSYL---GREGDKGLLPK-----STIV	489
Db	524	IVNREMFVSKKRY-----MEYANVYD--ILSWLNIAGIRVDNTTTSERKLHASTIK	576
Qy	490	QPAGS-----QYNTFYFDAAL--KKDYRLWYSTN-----TVGYREGGE---527	
Db	577	LHAQSDKGAYNRSMEEYQOTYADIMLVNKNKFGFNFLTANAGFSYEDHJTTGCMGGKLF	636
Qy	528	-----YTGYGSDDEEKRAFGENSPYKKHCNRSQI-----YEPVLKYYKKKRANNSHV	577
Db	637	TVPNLFSAY-----NFDPASGPGSQSHTHTRNNSVFSVELGYKSMLYLTLCGROEWASQ	691
Qy	578	SISADFCDYEMPASYSRTHRMNIOEY-----FSOIGDSGVHTALKP---621	
Db	692	LVNSDQTYEYPSGVSGV-----ISEMVSLEPKFTSFWMKRASFAEVGGPINYTLGTPT	746
Qy	622	-----ERANTWQFGFNTYKKGLLQDDTLGLKLKLVGYRSRDN	658
Db	747	VTDPMKGVINPISVYFPNFKAEQTSYELGTN-----LRLFNKNINIDA	792
Qy	659	--YIHNVYKGVNDLNDIGPSWSSTGLA-YTHOHRNFKDKVHKHGFLELNYD---YGR---	711
Db	793	TVTLTDYINOTF-----LSSMSPASGISGFYVQ-----AGKVRNKIGLESLGYNDRFGKVG	843
Qy	712	FFTNLSYAYOKS	723
Db	844	YATNLVYATANRN	855

RESULT 12

```

US-09-540-209B-7143
; Sequence 7143, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRAGILIS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 7143
; LENGTH: 833
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-7143

```

Query Match 3.5%; Score 172.5; DB 5; Length 833;
Best Local Similarity 19.8%; Pred. No. 7.2e-05;
Matches 187; Conservative 120; Mismatches 351; Indels 285; Gaps 44;

Qy	210	KKWLES	GASVGVLYGH	RRSRVAQNYRVGGQGHGNF	GAEBYLERRKQRYFVQEGAL	KPNS	269												
		: :	:	:	:	:													
Db	269	YTHLKT	GFTV-----	RA	FATNY-----	S	287												
		:	:	:	:	:													
Qy	270	DSGK	WERDLRQOQ	KYKPYKPNNOEL	OKYIEGHDKSWREN	LAPDYITPIDPSSL	KQOS	329											
		:	:	:	:	:	:												
Db	288	DN-----	DYKVF	PIIDL	ATNKKI-----	DERWVR	RNDAY-----	RS	320										
		:	:	:	:	:	:	:											
Qy	330	AGNL	FKLEYDGV	FNKYTAQ	FRDLNTKIGSR	KLIIRNYQFN	YGLSNPYTNL	NLTA	YNSG	389									
		:	:	:	:	:	:	:	:										
Db	321	GG--	IRLE	TCITN	KPYADY-----	LLAGI	ILSKN-----	DKDVQ	TGATMD	AVYGV	364								
		:	:	:	:	:	:	:	:	:									
Qy	390	RQK-----	YPKSG	FTGWL	KDXFETYN	NAKILDLN	NTA-TFRL	PRE	TELOTT	GLF	439								
		:	:	:	:	:	:	:	:	:									
Db	365	KMKSE	WIPSI	RYKKD	DLFDGL	SLSYGYN	SVNTF	TDTRARRY	NWLGES	VP	TSAGE	424							
		:	:	:	:	:	:	:	:	:	:								
Qy	440	NYFNE	XGKRF	PEEL	GLFDG	PDQDNG	LYSYLGR	FKGDKL	LPKOST	IV	PAG	SYFNT	499						
		:	:	:	:	:	:	:	:	:	:	:							
Db	425	GYYT	DSKIKR--	EWLG-----	NGNIS	Y-----	IDGHQ	SLTLN	HW	SAM	RRT	465							
		:	:	:	:	:	:	:	:	:	:	:							
Qy	500	FYF	DAALK	KDI	RLNYST	NTVGYR	FGGEGY	GYGSD	DEKFR----	AFGE----	NSPT	YVK	551						
		:	:	:	:	:	:	:	:	:	:	:							
Db	466	M-----	NKVR	DDENN	VPOOL	TKNITG-LGW	QIRDRN	NANV	FGKMY	KLYS	TYK	517							
		:	:	:	:	:	:	:	:	:	:	:							
Qy	552	HCNR	SCGI	YEP	VLKKY-----	GKRR	NNH	SVSIS	ADFG	DYF	MPFA-----	SYSR	TRH	MPN	601				
		:	:	:	:	:	:	:	:	:	:	:	:	:					
Db	518	-----	LDEY	TENAR	WEKVR	DHKT	NFCY	GAAAT	YYIL	PL	QA	KES	FES	HAY	RLPE	565			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:				
Qy	602	IQE	MYFSQ	IGDS	GV-----	HTALK	PER	ANT	WQ	FGFN-----	TYK	KG	LLK	QDD	TGL	KL	GV	YSR	655
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	566	SIEM	F-----	DGL	QO	RNP	OLPK	PESS	RLNL	GLUS	FTOT	GA	HQ	LSAD-----	608				
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Qy	656	IDN	IHN	VY	GKWD	NGD	IP	SSW	TSGL	AYTIO	HRNF	KDKV	HKH	GF	ELN	LDYD-GRE	FT	714	
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	609	-GN	IYR-YTT	D	FLKG-----	VS	LSNP-TT	GYENL-GK	VL	TG	VEA	AVR	NY	KDL	FHT	659			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Qy	715	NLS	YAK	SQ	TSPT	N	FSD	ASBS-----	PNN	ASKED	OL	KOG	YLS	RVS	AL	PRD	YGR	LEV	767
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	660	GAG	ETYQ	DI	TD	R													

RESULTS 13

```

US-09-540-209B-8474
; Sequence 8474, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540.209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 8474
; LENGTH: 702
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-8474

```

Query Match 3.5%; Score 171.5; DB 5; Length 702;
Best Local Similarity 19.6%; Pred. No. 6.7e-05;
Matches 149; Conservative 107; Mismatches 260; Indels 251; Gaps 39;
Qv 170 NLRITGVDDVVQGNNTYGLLLKIGITGRTNSTKGNMAAIGARKWLESGASGVLVGH8RRS 229

Db 95 SLNVQVSDAVK--HFAGVTVKDYGGIGGLKTVLSRLGAEH-----TAVG--YDGITIS 145
QY 230 VAQNYRVGGGQHGNGFAEYLER-----RKQYFVQEGALKFNSDSGKWERD 277
Db 146 DCQT-----GQIDIGFSLDNVRLSLNSGQSDNIQPARFFASAGIL-----N 189
QY 278 LQROQWKYKPYNNQELQKYTEGHDKSWRENLAPOYDITP---IDPSSLKQOOSAGNLF 334
Db 190 IOTLTQFQKDNR-----RTNLSASFKTGWSGLVNPSSLQLEQKLSRWK 231
QY 335 KLEVDG-----VFNKYTAQFRLNTKIGSRKLIINRYQENYGL--SLNPTNLNT 383
Db 232 VLSANGEMWADGHYPTLHYGEDNDLTSREKRNKTEVKNLRAEAGLFGNFSDEQWRK 291
QY 384 AAYNSGRQKYPKSGKGTGMLLKDFFETNNAKILDNL-----NTATFRLPTELOTTLGF 439
Db 292 AYYQSSRGLPNATY-----YDYSSOHLWDRNVFVOSQYKKEFSQWQFQSAKW 343
QY 440 NYFHNEYKGRNFPPEELGLFPDQDQNGLYSYLGRFKGDKGLLPQKSTIVOPAGSQYFNT 499
Db 344 NWSYQRY-----LD-PD-----YKSGSEG-----KTNSYYQOE 370
QY 500 FYFDA-ALKKDIYRLNYSTNT-----VGYREGGEVGYGYGSDDEFKRA 541
Db 371 YILSASALYRVLSNLSFSLSDASINRLNANLKDQFATPYRWLTAFAGYVND----- 424
QY 542 FGENSEPTYKKHCNRSQGIYEPVLK---KYGKKRANHSHVS--ISADFGDYFMPFAS----- 592
Db 425 -----WLTASASVLTATVINEEVRQGSAAANRRKLSPYVAS-----FKPFASEEPR 470
QY 593 -----YSTRHMPNIQEMYFSQIGDSGVHTALKPERANTWQFGNTYKKGLLKODDTLGLK 648
Db 471 IRLFYKDIFRLPSFNDLYGQVG-----NTNLKPESTTQYNLGL--TYSRST-----NE 517
QY 649 LVGYSR-RIDNYTHNVYKMWDLNGDIPS-----WYSSTGLAYTIQHRNFKDKVHKHGE 702
Db 518 LIPYVSTADAY-----YNKVKDIILPTKNLFIW-SWNLG-----KVDIKGID 562
QY 703 LELNYD---YGRFETNLS---YAYQKSTOPTNFSASESPNNAKEDQKOGYGLSRVSAL 757
Db 563 IAGNISLQPEKRLVNLSGNYTYQAL-----DMTEPGGKTYKQ-----QIAYT 606
QY 758 PRDYGRLEVQ--TRW--LGKLLGLGAMRYFGKSIRATAEERIDTNGTNGTNSFRLQK 813
Db 607 PRVSGSQAGIETPPWNLNLSYFLSKRYMLGNLNR-----ENRLDSYSDHSVSVRDLRI 662
QY 814 RSIKQETTLARQPLIFDFYAAEYPPKKNLIFRAEVKNLFDRRY 855
Db 663 RNVTSLTV-----EVLNLLDKNY 681

RESULT 14
US-09-584-501A-2
; Sequence 2, Application US/09584501A
; GENERAL INFORMATION:
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Potter, Andrew A.
; TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS OF
; TITLE OF INVENTION: PASTURELLA HAEMOLYTICA AND VACCINES CONTAINING THE SAME
; FILE REFERENCE: A34763 021645.0106
; CURRENT APPLICATION NUMBER: US/09/584,501A
; CURRENT FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: 08/753,759
; PRIOR FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: CA 2,164,274
; PRIOR FILING DATE: 1995-12-01
; PRIOR APPLICATION NUMBER: 60/008,569
; PRIOR FILING DATE: 1995-12-01
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

; LENGTH: 930
; TYPE: PRT
; ORGANISM: Pasteurella haemolytica
US-09-584-501A-2

Query Match 3.4%; Score 167; DB 5; Length 930;
Best Local Similarity 19.2%; Pred. No. 0.00023;
Matches 213; Conservative 142; Mismatches 375; Indels 380; Gaps 58;

QY 5 FRKPICFYLMGVTLVHYSYAEDAGRAGSEAOIQVLEDV-----HVAKRRVPKDKV 56
Db 8 FRISPVALTVLFSLSHYGAATENKKEENNDLAVLDEVIVTESHYAHERQNVGTGLKV 67
QY 57 FTDARAVSTRQDIFKFSSENLDNIVRSIPG-AFTQQDK--SSGIVSLNIRGDSGFRYNTM 113
Db 68 VKNYHEMSKQIIL-----GIRDLTRYDPGIVSVVEQGRGASSGYA--IRGVDK-NRYSLL 118
QY 114 VDGITO--TFYSTSTDRAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANL 171
Db 119 VDLGPOAHSYHTLGSAN--GGAIN---EIEYENIRSIELSKASSAEYSGAHGGAIGF 173
QY 172 RTLGVDDVVOGNTYGLLLK-GLTGTNSTKGNAMAAIARKWLESASGVLYVGHRSRV 230
Db 174 RTKDAQDIIEGQHWGLDSTSYASKNSHFLOSTAAAGEAGGFALVIATHRHGKETI 233
QY 231 AONYRVGGGQHGNGFAEYLERKQRYFVOEGALKFNSD-----SGKWERDLQROQW 283
Db 234 SEANKLKHNRITRTGTFENRYDTQIPHRMLLEDLLIVEDTCPTLDCPTPRARVKLNDRNF 293
QY 284 KYRPYKNYNQELQ-----KYIEGHDKSWRENLAPOYDITPIDPSSLKQOOSAGNLF 334
Db 294 PVRTFPEYTPPEERKQLEQIPYRTEQLSAQYTGKDRIAP-----NPLDYKS----- 339
QY 335 KLEVGVGFNKYTAQF-----RDLNTKIGSRKLIINRYQENYGLSLNPTY---NL 380
Db 340 ----NSVFMFGYHFNSSHYLGAILEDTKTRYDIRDQTPAY-----YTKDINL 385
QY 381 NLTAAYNSG-----RQYKPGSKFTGMLLKDFETNNNAKILDNNTATFRLPRET 431
Db 386 SLRNVYVGEQNLGLGVFKPRIPYGLR-----YSHVKFFD-----ER 422
QY 432 ELQTTLGFNFHNEYKGRNFPPEELGLFFDQDQNGLYSYLGR----- 474
Db 423 HKRRRLGFTYKYP-ENNRWLDSTIKLSAD--KODIELYSRLHRLHLCSDYPVVDKNCRPTL 479
QY 475 -----FKGDKGLLPQKSTIVQ-----PAGSOYFN-TFYFDAALKKDIYR--LNTSTN 518
Db 480 DKSWSMYRTERNYQEKHRYHLEFDFKALNAGOGVFNQTHKLNGLGDFRNSLMDHGM 539
QY 519 TVGYRFGGEYTYGSDDEKRAFGENSPYKHH-----CNRSCGIY---EPVLKK 566
Db 540 TAQYTKGG-YTSYRG-----RGRLDNPIYRRDPRSIETVSLCNTRGDTLNCPE--- 588
QY 567 YGKKRANHSHVSIADFGDYFMPFASYSRTHRMPNIEEMYFSQIGDSGVHTALKPERANT 626
Db 589 -RKIKGDSHEFVS-----FRLDVISEYVDLGLGVRFQDHR--- 621
QY 627 WQFGENTYKKGLLKQDDTLGLKLVGYSRIDNYTHNVYKMWDLNGDI---PSWVSSTG 682
Db 622 -----FKSDDPDWLTSRT-YRN-----WSWNGGTLTKPTEFVS--- 652
QY 683 LAYTIQHRNFKDKVHKHGFLELNYD-YGRFFTNLSYAYOKSTOPTNFSASESPNNAK 741
Db 653 LSVRI-----SNGFRVPAPFELYGK-----RDHIGLKD-NEVYQRAQR 689
QY 742 EDQLKQGYLSRVSALPRDYCRLEVGTRWLGKNTLGLGAMRYFG--KSTRATAEERYID 798
Db 690 SHOLE-----PEKSTNHEIGVSFKQFGYLD--VSIFRNNYKNMIATACKRIIQ 736
QY 799 GTNGGNTSNFRLQKRSIKQETTLARQPLIFD-----FYA--AYEPKKNLIFRA 845
Db 737 KSH--CFYNTHNQDVALNGINLVAK----FDLHGILSMLPDGIFYSSVAVN-----RV 783

THIS PAGE BLANK (USPTO)

```
Db 128 VSFQEGDDPATVINIRGLQDFGRVAVVVGARQNYQT-----GHNANGSFF---LPEL 180
Qy 146 IAGLDVVKVSGSAGINSLAGSANTLRLLVDDVVQGNNTYGLLKLGLTGNSTKGNAMA 205
Db 181 IGGVDVVRGPTANIYGSAGLGLVSTFKDINDVLRPGERGVDSLGSYGNRRGLGSV 240
Qy 206 AIGARKWLESCAGVLYGHSSRVAQNYRVCGGQHGAEYELRRKQRYFVQEGAL 265
Db 241 FGGVRAIPDVDFGAVY-----RTQNGYK-DGNGTEIGTNGQ-----VEAGLM 284
Qy 266 KFNDSKGWERDLQROQWKYPKYNQELQKYEIGHDKSWRENLAPOYDITPIDPSSL 325
Db 285 KLT-----VRP-----ALGHEV----- 296
Qy 326 KQOSAGNLFKLEYGVFNKYTAQFRDLNTKGSRKLIINRNQYFNGYSLNPTNLNTAA 385
Db 297 -----KFGAVEQDYQYDYG-----QFNRG-----PTTQAL-IA 324
Qy 386 YNSGRQKYPKSK-FTG-----WGLLKD-----FETYNNAKILDLNNTATFRLPRETEL 433
Db 325 LNRSSVYASDAKNYSGVTWNYSLPSDLNLFDMHMSVYGNR--TDNQTKTYHYGTTPSA 382
Qy 434 QTTLGF-----NYPHNEVG--KNRFPPEELGLFFDGDODNGLYSYLGRFGKD- 478
Db 383 YCNGGFGNNVSGVCDKRGYVLTNYGYDANNTR-----FNVGWRNALTWGVDAFQDDV 437
Qy 479 -----KGLLPKOKSTIVOPAG-----SOYENTFYFDA--ALKKDIYRLNY-STNT 519
Db 438 ITTDSRG-----NSNITTPSGIRTVSGGFLQLKQNYST-WLEAVSAIRYDRYDLDSGKTST 492
Qy 520 VGYRFGYEYTYGSDDEFRKAFGENSPYKHKHCRSCGIEPVLYKKGKRRANNHVSVI 579
Db 493 GDRFSPKIT-----LGV-TPV----- 508
Qy 580 SADFGDYFMPFASYSRTHRPNIOEMVFSQIGDSGVHTA-----SGAHATGGGPAFFVCPDGTAGLFCFLP 558
Db 509 -----PGFQPVVSAEGRASITETVI-----SGAHATGGGPAFFVCPDGTAGLFCFLP 558
Qy 619 ---LKPERANTWQGFNTYKGLLKQDDTLGLKLGVGRSRIDNYIHNYGKMWDLNGDIP 675
Db 559 NPNLRPEVGKKEVINLKYDNIESANDSPRGKINLFRNDVSDVI-----DLVASAP 610
Qy 676 SWVSTG-LATYIGHRFKDKVHKHGELELNYDYGRFTNLSTAY-QKSTQPTNFSAS 733
Db 611 VAVPPGFSFQYQYQNIAN-ARIRGFEAETMYDAGDWFIGVAGHYIQGKNVATNIGLAT 669
Qy 734 ESPNASKEDQKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGLGAMRYFGKSIRATAE 793
Db 670 ITP-----RKVVTGGVRLLDRLILITAQ 693
Qy 794 ERYIDGNGGNTSNRQLGKRSIKQETLARQPLIFDYAAEYEPKKNLIFRAEYKNLFD- 852
Db 694 WASYGPNDVPAGYLPATGYE-----LVNLTYTNATRDIVLSASIDNLLNQ 740
Qy 853 ---RRYIDPLDAGNDAATQ 868
Db 741 YRYPYAIIP-GSSTGDTTQ 757

RESULT 15
Q9KLM6 PRELIMINARY; PRT; 784 AA.
AC Q9KLM6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TONB RECEPTOR-RELATED PROTEIN.
GN VCA0625.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX STRAIN-EL TOR N16961 / SEROTYPE O1;
RA MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004392; AAF96526.1; .
DR TIGR; VCA0625; .
DR InterPro; IPR000531; TonB_boxC.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 784 AA; 87575 MW; 605DCFF12B0CBB10 CRC64;

Query Match 6.3%; Score 309.5; DB 16; Length 784;
Best Local Similarity 21.2%; Pred. No. 5.2e-09;
Matches 195; Conservative 122; Mismatches 343; Indels 261; Gaps 43;

Qy 6 RLKPICFYLMGVTLHYSYAEDAGRAGSEAQIOVLVDHVHAKRVPKDKVFTDARAVST 65
Db 36 QLKPL-FTLLPVLSSVVQAQE---NTEQAVDETVTVH-----QQSILTDQR---T 79
Qy 66 RODIFKSSNDLNVRSIPGAFTQDKSGGIVS-----LNIRGDSGFGRVN 111
Db 80 RSDLOK-----VRGIANA---DIFSGITSVOSNMHNEAGALDIGRGVQGEGRVP 127
Qy 112 TMVDGITQTFYSTDAGRAGSSQFQASVDSNFTAGLDVVVKG-----FSGSAGINS 164
Db 128 IFIDGSLQ---STHTRGYQGVSDR--TYIDTLLSSLTVNKGATIESSPYASGAVG--- 179
Qy 165 LAGSANLRTLGVDVQGNNTYGLLLKG-----LTGTNSTKGN-AMAAIGARKWL 213
Db 180 --GVVNATTGLIKDQDAFVYVKARANNHNTPDVSGDYSEQGYALDERGHSF 237
Qy 214 ESGA-----SVGVLYGHSRVSVAQNYR-----VGGGGQHIGNFGCAEYLE 252
Db 238 KHGSLMLGLGVAQASFTNLVAYSRRKKNHFGAGKGYEYQEPVVGQGVVNTSFE--- 294
Qy 253 RRKORYFVQEGALKFNDSGKWERDLQROQWKYPKYNQELQKYEIGHDKSWRENLA 312
Db 295 -----SDSWLFKIASDTG---TAHNADFNRYRHAQKAGEVLWAYWYKSSDEWGNPY 343
Qy 313 POYDITPIDPSSLKQOSAGNLFKLEYGVFNKYTAQFRDLNTKIGSRKLIINRNQYFNGL 372
Db 344 PD-----GKDRMPQWGLGT-----AKVNTYSA-----NYYQPDH--- 373
Qy 373 SLNPTNLNLTAAYNSGRQKYPKSGKFTG-WGLLKDFETYNNAKILDLNNTATFRLPRET 431
Db 374 ---PWLNLNANFWYTEADL-----AOYNGLMALGTNABOYFHAY---HNRSGSLTNET 422
Qy 432 EL-QTTLGFNYFHNEYGKNRPEELGLFFDGDODNGLYSYLGRFKGDKLLPQKSTIVQ 490
Db 423 LLTQMPVRLNYGLAQNERLSPEE-----DGQT-----RE-----TKTVTS 458
Qy 491 PAGSOYFNWTFYDAALKKDIYRLNYSNTNVTGYRFGGEYTYGSDDEPKRAFGENSPYK 550
Db 459 RHGKRRTAQNLFANADIDYSPLRVQLGLNLHNAK-----STDYQTKQOL-----DYK 504
Qy 551 KHCNRSCGIIPEPVLYKKGKRRANNHVSISADFGDYFMP-----FASYSTRHMPNIOEMY 606
Db 505 E-----KDLLESTYALTPSTQFLKLSRTYRMPSLYETT 540
Qy 607 FS-QIGDSGVHTALKPERANTWQGFNTYKGLLKQDDTLGLKLVGRSRIDNYIHNVYG 665
Db 541 LSNVEFSYNPNPKPEQAWNNEVGQVMASNSVLQDRLNLSVSYFNSLKDFISG--G 598
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:53:15 ; Search time 62.3 seconds
(without alignments)
1642.038 Million cell updates/sec

Title: US-09-762-926-6
Perfect score: 4894
Sequence: 1 MRSSRLPFCFLMGVTLV.....SVLTNFAQRFLITMSYKF 921

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4894	100.0	921	21	AAY69382 A BASB024 Outer me
2	4779.5	97.7	922	21	AAY69381 A BASB024 Outer me
3	4769.5	97.5	922	21	AAY69380 A BASB024 Outer me
4	4707.5	96.2	922	20	AA138940 N. gonorrhoeae ant
5	4703	96.1	888	20	AA138937 Neisseria meningit
6	4696.5	96.0	922	20	AA138939 Neisseria gonorrh
7	4348.5	88.9	871	20	AA138938 N. meningitidis st
8	2588.5	52.9	947	21	AA144428 M. catarrhalis (AT
9	2505	51.2	913	21	AA194671 Haemophilus antige
10	2483.5	50.7	918	21	AA194672 Haemophilus antige
11	1996	40.8	393	20	AAY38936 Neisseria meningit

12	898	18.3	753	22	ABB52934 Escherichia coli p
13	898	18.3	753	22	ABB52968 Escherichia coli p
14	273.5	5.6	915	20	AAV07476 N.gonorrhoeae trans
15	273.5	5.6	915	21	AAV51770 N. gonorrhoeae str
16	273.5	5.6	915	21	AAV80374 N. gonorrhoeae str
17	269	5.5	908	21	AAV51768 N.meningitidis str
18	269	5.5	908	21	AAV80372 N.meningitidis st
19	267.5	5.5	791	17	AAV85566 N.meningitidis se
20	264.5	5.4	790	17	AAV95565 N.meningitidis se
21	264	5.4	971	21	AAV21223 Haemophilus somnus
22	256.5	5.2	911	21	AAV51769 N.meningitidis str
23	256.5	5.2	911	21	AAV80373 N.meningitidis st
24	255.5	5.2	915	22	AAE12018 Neisseria meningit
25	253.5	5.2	909	20	AAV07477 N.meningitidis tra
26	252.5	5.2	884	14	AAV34402 Sequence of low mo
27	251.5	5.1	908	15	AAV48220 N.meningitidis IM2
28	249.5	5.1	790	17	AAV95568 N. gonorrhoeae B h
29	248.5	5.1	912	16	AAV96969 Transferrin recept
30	246	5.0	792	17	AAV95567 N.meningitidis se
31	230.5	4.7	912	18	AAV08959 Amino acid sequenc
32	230.5	4.7	912	18	AAV08961 H. influenzae str
33	230.5	4.7	912	19	AAV53044 H. influenzae str
34	230.5	4.7	912	19	AAV53046 H. influenzae str
35	230.5	4.7	912	21	AAV51689 H. influenzae type
36	230.5	4.7	912	21	AAV51691 H. influenzae type
37	230.5	4.7	912	21	AAV80355 H. influenzae type
38	230.5	4.7	912	21	AAV80357 H. influenzae type
39	230	4.7	913	16	AAV77884 Bacterial transfer
40	228.5	4.7	887	14	AAV34403 Sequence of low mo
41	228.5	4.7	912	16	AAV77886 Bacterial transfer
42	228	4.7	715	17	AAV01461 NTHI HXUC protein.
43	228	4.7	911	16	AAV77892 Bacterial transfer
44	228	4.7	911	18	AAV08966 Amino acid sequenc
45	228	4.7	911	19	AAV54124 H. influenzae str

ALIGNMENTS

RESULT 1
AAY69382
ID AAY69382 standard; Protein; 921 AA.
AC AAY69382;
XX
XX
DT 19-JUN-2000 (first entry)
XX
XX
DE A BASB024 outer membrane protein of N. meningitidis.
XX
KW BASB024; outer membrane protein; N. meningitidis infection;
KW Bacteremia; meningitis.
XX
OS Neisseria meningitidis.
XX
XX
PN WO200011182-A1.
XX
PD 02-MAR-2000.
XX
PF 13-AUG-1999; 99WO-EP05989.
XX
PR 18-AUG-1998; 98GB-0018004.
XX
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Thonard J;
XX
XX
DR WPI; 2000-224702/19.
N-PSDB; AA261582.
XX
XX Novel polypeptides derived from the products of the BASB024 gene of
PT Neisseria meningitidis, useful for inducing an immune response and
PT producing antibodies useful for treating meningitis -
XX

PS Claim 3; Page 91-95; 103pp; English.

XX The present sequence represents a BASB024 outer membrane protein of

CC Neisseria meningitidis. The BASB024 polynucleotide sequence was

CC first identified in the Incyte Pathoseq database containing

CC unfinished genomic DNA sequence of N. meningitidis. BASB024

CC polypeptides and polynucleotides are useful for generating an

CC immune response in an animal. Antibodies specific BASB024 polypeptides

CC are useful for treating N. meningitidis infection, which causes

CC bacteremia and meningitis.

XX

SQ Sequence 921 AA;

Query Match 100.0%; Score 4894; DB 21; Length 921;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 921; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSSFRLKPCIFYLMGVTLHYHYSAEDAGRAGSEAOIQVLEDVHVAKRVPKDKKVFETDA 60

DB 1 mrssfrlkipcifylmgvtlhyhsyaedagragseaqiqlvledvhvakrvpkdkkvvftda 60

QY 61 RAVSTRODIFKSSNLDNIVRSIPGATQODKSSGIYSLNIRGDSGFRVNTWVDGITQT 120

DB 61 ravstrdqifkssnldnivrsipgattqodkssgivslnirgdsgrvntwvdmvgitqt 120

QY 121 FYSTSTADGRAGSSQFASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLGVDVV 180

DB 121 fyststdagragssqfagsvdsnfiaqldvvkgsfsgsaginslagsanlrtlvgddvv 180

QY 181 QGNNTYGLLLKGLTGTNSTGNMAAIGARKWLESASGVLYGHSRVSVAQNYRVGGGG 240

DB 181 qgnntygllllkglgtgstnsgnmaaiarkwlesasgvlyghsrvsvaqnyrvgggg 240

QY 241 QHIGNFAEYLERKQRYFVQEGALKFNSDSGKWERDLQOQWKYKPYKNYNQLOKYYI 300

DB 241 qhignfaeylerkrqryfveqegalkfnsdsgkwerdlrqoqwkypkyknynnqelkyyi 300

QY 301 EGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYYTAOFRDLNKTIGSRK 360

DB 301 eghdkswrenlapoyditiPIDPSSLKQOSAGNLFKLEYDGVFNKYYTAQFRDLNKTIGSRK 360

QY 361 IINRNYQFVGLSLNPYTNLNTAAYNSGRQKYPKSGKFTGWGLLKDFEYNNNAKILDLN 420

DB 361 iinrnyqfnyglslnpytlnlntaaynsgrqkypksgkftgwglldkfetynnakildln 420

QY 421 NTATFRLPRETELQTLGCFNYFHNKRFPEELGLFFDGPQDNGLYSVLGRFKDGK 480

DB 421 ntatfrlpretelqtltlgcfnyfhnekrfpeelglffdgpdnglysvlgrfkdgk 480

QY 481 LLPQKSTIVOPAGSQYFNTFYFDAALKKDIYRLNYSNTVGYRFGGEYTYGYSDDFEKR 540

DB 481 llpqkstivopagsqyftntfyfdaalkkdiyrlnysntvgyrfggetytygsddfekr 540

QY 541 AFGNSPTYKHKNRSGIPEVLKYGKRRANNHVSISADFGDYFMPFASYSRTHRMP 600

DB 541 afgnsptykhnrcsgipevlkygkrrannhvsisadfgdyfmpfasystrhrmp 600

QY 601 NIOEMYSQIDSGVHTALPERANTWQFCNTYKGLLKDDPFLGLKLYGYSRIDNYI 660

DB 601 niqemyfsqidsgvhtalperantwqfntkykglldpflglklygyrsridnyi 660

QY 661 HNVYCKWMDLNGDIPSWSSVTGLAYTTQHRNFKDKVHKHGFELNLYDYGRRFTNLISYAY 720

DB 661 hnvycgkwmldngdipswssvtglayttqhrnfkdvkhkhgfelnlydygrfntlnisyay 720

QY 721 OKSTQPTNFSDASSPNNAKEDOLKOGYGLSRVSALPRDYGRLVETRWLGNKLTIGGA 780

DB 721 okstqptnfdsasespnnaakedolkogyglsrvsalprdygrlevetrwlgnkltlga 780

QY 781 MRYFGKSTRATAERYIDGTNGGNTSNFROLGKRISIKQETLAROPLIFFDYAAYEPKKN 840

DB 781 mryfgkstrataeryidgtngngntsnfrlgrksikqetlarqpliffdyayepkkn 840

QY 841 LIFRAEVKNLFDTRYIDPLDAGNDAAATQRYYSFDPKDKDEDTCTNADKTLNCKYGGTS 900

DB 841 lifraevknlfdrdyidpldagndaatqrryysfdpdkdktcndadktnclngkygts 900

QY 901 KSVLTNFARGRTFLITMSYKF 921

DB 901 ksvltnfargrtflitmsykf 921

RESULT 2

AAV69381

ID AAY69381 standard; Protein; 922 AA.

XX

AC AAY69381;

XX

DT 19-JUN-2000 (first entry)

XX

DE A BASB024 outer membrane protein of N. meningitidis.

XX

KW BASB024; outer membrane protein; N. meningitidis infection;

KW bacteremia; meningitis.

XX

OS Neisseria meningitidis.

XX

PN WO200011182-A1.

XX

PD 02-MAR-2000.

XX

PF 13-AUG-1999; 99WO-EP05989.

XX

PR 18-AUG-1998; 98GB-0018004.

XX

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Thonnard J;

XX

DR WPI; 2000-224702/19.

DR N-PSDB; AA261581.

XX

PT Novel polypeptides derived from the products of the BASB024 gene of

PT Neisseria meningitidis, useful for inducing an immune response and

PT producing antibodies useful for treating meningitis -

XX

PS Claim 3; Page 88-91; 103pp; English.

XX

CC The present sequence represents a BASB024 outer membrane protein of

CC Neisseria meningitidis. The BASB024 polynucleotide sequence was

CC first identified in the Incyte Pathoseq database containing

CC unfinished genomic DNA sequence of N. meningitidis. BASB024

CC polypeptides and polynucleotides are useful for generating an

CC immune response in an animal. Antibodies specific BASB024 polypeptides

CC are useful for treating N. meningitidis infection, which causes

CC bacteremia and meningitis.

XX

SQ Sequence 922 AA;

Query Match 97.7%; Score 4779.5; DB 21; Length 922;

Best Local Similarity 97.9%; Pred. No. 0;

Matches 903; Conservative 7; Mismatches 11; Indels 1; Gaps 1;

QY 1 MRSSFRLKPCIFYLMGVTLHYHYSAEDAGRAGSEAOIQVLEDVHVAKRVPKDKKVFETDA 60

DB 1 mrssfrlkipcifylmgvtlhyhsyaedagragseaqiqlvledvhvakrvpkdkkvvftda 60

QY 61 RAVSTRODIFKSSNLDNIVRSIPGATQODKSSGIYSLNIRGDSGFRVNTWVDGITQT 120

DB 61 ravstrdqifkssnldnivrsipgattqodkssgivslnirgdsgrvntwvdmvgitqt 120

QY 121 FYSTSTADGRAGSSQFASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLGVDVV 180

DB 121 fyststdagragssqfagsvdsnfiaqldvvkgsfsgsaginslagsanlrtlvgddvv 180


```
QY 181 QGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESASGVLYGHSRRSVAQNTRVGGG 240
DB 181 qgnntyglllkgltgtnstkgnaaaigarkwlesasgvlyghsrrsvaqntrvvggg 240
QY 241 QHIGNFGAEYLERRKORYFVQEGALKFNDSGKWERDLQROQWKYPKYNKN--QELQKY 299
DB 241 qhignfgaeylerrkoryfvqegalkfnsgkwerdfqrpywktkwyqyndpqelqky 300
QY 300 IEGHDKSWRENLAPOYDITPIDPSLSKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 359
DB 301 ieghdkswrenlapqyditpidpslskqosagnlfkleydgvfnkytaqfrdlntkigsr 360
QY 360 KIINRYOFNYGLSLNPTNLNTAAAYNSGRQKYPKSGKFTGWLGLLKDFEYNNAKIIDL 419
DB 361 kilnrnyofnyglslnsyanlnltaaynsgrqkypksgkftgwglldkfetynnakildl 420
QY 420 NNTATFRLPRETELQTTLGFNFYHNEYCKNRPPEELGLFFDGPDDNGLYSLGREFGDK 479
DB 421 nntatfrlpretelqttlgfnfyhneygknrfpeelglffdgppdnglyslgrfkgd 480
QY 480 GLLPKQSTIVQAGSOYFNTFYDAALKKDIYRLNYSNTNTVGYRFGGEYTYGYGSDDEFK 539
DB 481 glpqsstivqagsyftntfydaalkkdiyrlnystntvgyrfggytygygsddefk 540
QY 540 RAFGENSPYKKHCNRCGIIYEPVLKKYKKRANNSHVSISADFGDYPMPFASYSRTHRM 599
DB 541 rafgensptykkhcnrcgiiyepvlkkygkkrannshvsisadfgdyfmpfasyrthrm 600
QY 600 PNIQEMFSQLGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 659
DB 601 pniqemyfsqldsgvhtalkperantwqgfntykkglkqddtlglklvgysrridny 660
QY 660 IHNYGKWDNLNGDIPSVSSTGLAYTIOHRNFKDKVHKHGFLELNYDYGRFFTNLSYA 719
DB 661 ihnygkwdnlngdipsvsstglaytiohrnfkdvkhkhgflelnydygrfftnlsya 720
QY 720 YOKSTQPNFSDASEPNNAKEDQLQKGYGLSRVSALPRDYGRLVGTWRLGNKLILGG 779
DB 721 yqkstqpnfstdasepnnaakedqlqkgyglsrvsalprdygrlevgtwrlgnkltlg 780
QY 780 AMRYFGKSIRATABERYIDGYNNGNTSNFRQLGKRSIKQTETLARQPLIFDFYAAEPKK 839
DB 781 amryfgksirateaeryidgntngntsnvrqlgkrsikqtetlarqplifafyaaepkk 840
QY 840 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLCKNGKYGT 899
DB 841 nlifraevknlfdriryidpldagndaatqryysfdpkdkdedvtcnadkctlcngkygt 900
QY 900 SKSVLTNPFARGTRTFLTMSYKF 921
DB 901 sksvltnfargtrtfltmsykf 922

RESULT 3
AAV69380
ID AAV69380 standard; Protein; 922 AA.
XX
AC AAV69380;
XX
XX
DT 19-JUN-2000 (first entry)
DE
DE A BASB024 outer membrane protein of N. meningitidis.
KW BASB024; outer membrane protein; N. meningitidis infection;
KW bacteremia; meningitis.
OS Neisseria meningitidis.
XX
XX WO200011182-A1.
XX
PD 02-MAR-2000.
XX
```

```
PF 13-AUG-1999; 99WO-EP05989.
XX
PR 18-AUG-1998; 98GB-0018004.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2000-224702/19.
DR N-PSDB; AAZ61580.
```

Novel polypeptides derived from the products of the BASB024 gene of *Neisseria meningitidis*, useful for inducing an immune response and producing antibodies useful for treating meningitis -

Claim 5; Page 83-86; 103pp; English.

The present sequence represents a BASB024 outer membrane protein of *Neisseria meningitidis*. The BASB024 polynucleotide sequence was first identified in the Incyte Pathoseq database containing unfinished genomic DNA sequence of *N. meningitidis*. BASB024 polypeptides and polynucleotides are useful for generating an immune response in an animal. Antibodies specific BASB024 polypeptides are useful for treating *N. meningitidis* infection, which causes bacteremia and meningitis.

Sequence 922 AA;

Query Match 97.5%; Score 4769.5; DB 21; Length 922;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 901; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

```
QY 1 MRSSFRLKPICTCYLMGVTLHYHYAEADAGRAGSEAIQVLEDVHVHAKRVPKDKVFTDA 60
DB 1 mrssfrikpicfyimgvmlhyhsyaedagragseaqiqlvledvhvakrvpkdkkvfda 60
QY 61 RAVSTRQDIFKSSENLDNIVRSIFCAFTQDKSSGIVSLNIRGDSGFRVNTWVDGIQT 120
DB 61 ravstrqdifksenldnivrsipgafqtqdkssgivslnirgdsgrvntwvmdgitqt 120
QY 121 FYSTSDAGRAGGSSOFGASVDSNFIAGLDVVKGFSGSAGINSLAGSANLRTLGVDDVV 180
DB 121 fyststdagraggsqfsgasvdsnfiagldvvkgsfsgsaginslagsanrltlygvddv 180
QY 181 QGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESASGVLYGHSRRSVAQNTRVGGG 240
DB 181 qgnntyglllkgltgtnstkgnaaaigarkwlesasgvlyghsrrsvaqntrvvggg 240
QY 241 QHIGNFGAEYLERRKORYFVQEGALKFNDSGKWERDLQROQWKYPKYNKN--QELQKY 299
DB 241 qhignfgaeylerrkoryfvqegalkfnsgkwerdfqrpywktkwyqyndpqelqky 300
QY 300 IEGHDKSWRENLAPOYDITPIDPSLSKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 359
DB 301 ieghdkswrenlapqyditpidpslskqosagnlfkleydgvfnkytaqfrdlntkigsr 360
QY 360 KIINRYOFNYGLSLNPTNLNTAAAYNSGRQKYPKSGKFTGWLGLLKDFEYNNAKIIDL 419
DB 361 kilnrnyofnyglslnsyanlnltaaynsgrqkypksgkftgwglldkfetynnakildl 420
QY 420 NNTATFRLPRETELQTTLGFNFYHNEYCKNRPPEELGLFFDGPDDNGLYSLGREFGDK 479
DB 421 nntatfrlpretelqttlgfnfyhneygknrfpeelglffdgppdnglyslgrfkgd 480
QY 480 GLLPKQSTIVQAGSOYFNTFYDAALKKDIYRLNYSNTNTVGYRFGGEYTYGYGSDDEFK 539
DB 481 glpqsstivqagsyftntfydaalkkdiyrlnystntvgyrfggytygygsddefk 540
QY 540 RAFGENSPYKKHCNRCGIIYEPVLKKYKKRANNSHVSISADFGDYPMPFASYSRTHRM 599
DB 541 rafgensptykkhcnrcgiiyepvlkkygkkrannshvsisadfgdyfmpfasyrthrm 600
```

Qy 600 PNIQMYFSQIGDSGVHTALKPERANTWQFGNTYKKGLLKQDDTLGLKLGYRSRIDNY 659
Db 601 pniqemyfsqigdsghvhtalkperantwqfgntyykkgllkqddtlglklgyrsridny 660
Qy 660 IHNVYKWKWDLNGDIPSWSVSTGLAYTIQHRNFKDKVHKHGFELNLDYGRFTNLSYA 719
Db 661 ihnvvgkwdlmgpwsvsstglaytiqhrnfkdvkhkgfelnldygrfftnlsya 720
Qy 720 YOKSTOPTNFSDESPPNASKEDQKQGYGLSRVSALPRDYGRLEVGTRWLGKNTLGG 779
Db 721 yqkstqptnfsdaespnnaskedqlkqgyglrsvalprdygrlvgtrwlgkntlgg 780
Qy 780 AMRYFGKSIRATAERYIDGTNGNTSNFRLGKRSIKQTETLARQPLIFDYAAYEPKK 839
Db 781 amryfgksirataeeryidgtngntsnvrqlgkrsikqtetlarqplifdyaayepkk 840
Qy 840 NLIFRAEVKNLFDRRYIDPLDAGNDAATQRYSSFDPKDDEYTCNADKTLGCKYGGT 899
Db 841 nlifraevknlfdrdyidpldagndaatqryssfdpkddeytcnadtclngkyggt 900
Qy 900 SKSVLTNFARGRTFLITMSYKF 921
Db 901 sksvltnfargrtflitmsyxf 922

RESULT 4
AAV38940
ID AAV38940 standard; Protein; 922 AA.
XX
AC AAV38940;
XX
DT 08-OCT-1999 (first entry)
XX
DE N. gonorrhoeae antigen encoded by a variant ORF133.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
XX
PR 06-NOV-1997; 97GB-0023516.
XX
PR 14-NOV-1997; 97GB-0024190.
XX
PR 18-NOV-1997; 97GB-0024386.
XX
PR 27-NOV-1997; 97GB-0025158.
XX
PR 10-DEC-1997; 97GB-0026147.
XX
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masiagnani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WPI: 1999-327407/27.
XX
DR N-PSDB: AA212354.
XX
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
PS Claim 4; Page 480; 524pp; English.
XX
XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids

CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 922 AA;

Query Match 96.2%; Score 4707.5; DB 20; Length 922;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 885; Conservative 15; Mismatches 21; Indels 1; Gaps 1;

Qy 1 MRSSFRLKPTCFYLMGVTLHYHYAEADAGRAGSEAOQVLEDVHVAKRVPKDKKVTDA 60
Db 1 mrsfrlkipcfylmgvmlhyhsyaedagragseaqvledvvhkarkrpkdkkvtdda 60
Qy 61 RAVSTRDIDFKSSNLDNIVRSIPCAFTQODKSSGIVSLNIRGDSGFRVNTWVDGTTQT 120
Db 61 ravstrdgvfksngnldnivrslpcaftqodkssgivslnirgdsgrvntwvdgttqt 120
Qy 121 FYSTDAGRAGSSQFGASVDSNFIAGLDVYKGSFSGSAGINSAGSANTLRTIGVDDV 180
Db 121 fystdtagragssqfgasvdsnfiagldvvykgsfsgsaginsagsanlrtlvgddv 180
Qy 181 QGNNTYGLLLKGLTGTNSTKGNMAAATCARKWLESASVGLYCHSRSSVAQNYRVGGG 240
Db 181 qgnntyglillkgltgtnstkgnmaaaigarkwlesgasvglvghsrgrvaqnyrvggg 240
Qy 241 QHIGNFGAEYLERRKQRYFVOEGALKFNSDGSQKWERDLQROQWKYKPKYNN-OELQKY 299
Db 241 qhignfgeeylerkqyfvqegglkfnagskwerdlrqywktykkyedqelqky 300
Qy 300 IEHDKSWRENLAPOYDITPIDPSSLKQSQAGNLFKLEYDGVFNKYTAQPRDLNTKIGSR 359
Db 301 ieehdkswrenlapoyditiidpsglkqsgagnlfkleydgvfnkytaqfrdlntigrs 360
Qy 360 KIINRNYQFNYSGLNPNYTNLNTAAVNSGRQKYPKSKETGWLKDKFETYNNAKILDL 419
Db 361 kiinrnyqfnyslnpytnlnitaaaynsgrkypkagktgwlkdfetynnakilld 420
Qy 420 NNTATFLPRETELEOTTLGFNFYHNEYKKNRFPPELGFDPDQDNGLSYGLRFRGDK 479
Db 421 nntatflpretelqtlgfnfneygknrfpeelgffdgpdqnglsyglrfgdk 480
Qy 480 GLLPQKSTIVQAGSQYFNTFYFDDAALKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 539
Db 481 glipqkstivqagsyfnfyfdaalkdiyrlnystnainyrfggeytgyyseneff 540
Qy 540 RAFGENSPYKKHCRNRCGIEPVLLKYGKRRANNHVSISADPGDYFEMPASYSRTHRM 599
Db 541 rafgenspaykehcdpscglyepvllkygkrrannhsvsasadfyfmpfagsyrthrm 600
Qy 600 PNIQEMYFSQIGDSGVHTALKPERANTWQFGNTYKKGLLKQDDTLGLKLGYRSRIDNY 659
Db 601 pniqemyfsqigdsghvhtalkperantwqfgntyykkgllkqddtlglklgyrsridny 660
Qy 660 IHNVYKWKWDLNGDIPSWSVSTGLAYTIQHRNFKDKVHKHGFELNLDYGRFTNLSYA 719
Db 661 ihnvvgkwdlmgpwsvsstglaytiqhrnfkdvkhkgfelnldygrfftnlsya 720
Qy 720 YOKSTOPTNFSDESPPNASKEDQKQGYGLSRVSALPRDYGRLEVGTRWLGKNTLGG 779
Db 721 yqkstqptnfsdaespnnaskedqlkqgyglrsvalprdygrlvgtrwlgkntlgg 780
Qy 780 AMRYFGKSIRATAERYIDGTNGNTSNFRLGKRSIKQTETLARQPLIFDYAAYEPKK 839
Db 781 amryfgksirataeeryidgtngntsnvrqlgkrsikqtetlarqplifdyaayepkk 840
Qy 840 NLIFRAEVKNLFDRRYIDPLDAGNDAATQRYSSFDPKDDEYTCNADKTLGCKYGGT 899
Db 841 nlifraevknlfdrdyidpldagndaatqryssfdpkddeytcnadtclngkyggt 900
Qy 900 SKSVLTNFARGRTFLITMSYKF 921
Db 901 sksvltnfargrtflitmsyxf 922

```
RESULT 5
AAV38937
ID AAY38937 standard; Protein; 888 AA.
XX
AC AAY38937;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis antigen encoded by a partial ORF133.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria meningitidis.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WP1; 1999-327407/27.
DR N-PSDB; AA212352.
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
PS Claim 4; Page 474; 524pp; English.
XX
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 888 AA;

Query Match 96.1%; Score 4703; DB 20; Length 888;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 885; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 EAQIQVLEDVHVAKRVPKDKVFTDARAVSTRQDIFKSSSENLDNIVRSPGAFTOQDKS 93
D 1 eaqiqlvledvhvakrvpdkdkvftdaravstrqdifksssenldnivrspgaf tqdks 60

QY 94 SGIVSLNIRGDSGFRVNTMWVDGTTQTFYSTSTDTAGRAGSSQFGASVDNFTAGLDVVK 153
D 61 sgivslnirgdsgrvntmvdgttqtfyststddagragssqfgasvdsnfia gldvvk 120

QY 154 GSFSGSAGINSLAGSANLRTGLVDVDDVVGNNVTYGLLLKGLTGNTSKGNAMAAIGARKWL 213
D 121 gsfsgsaginslagsanlrtlglvddvddvvgnnvtyglllkgltgntskgnamaaigarkwl 180

QY 214 ESGASGVLYGHSSRRSVAQNRVYGGGGQHTGNFGAEYLERRKORYFVQEGALKFNSDSGK 273
D 181 esgasgvlyghssrrsvaqnrvyggggqhlgtnfgaeylerkroryfvqegalkfn sds gk 240
```

```
QY 274 WERDLQRQWKYKPYKNYNNOELQKYIEGHDKSWRENLAPOYDITPDPSLKKQOSAGNL 333
D 241 werdlqrqwkypkynynnqelqkyieehdkswrenl xpyqditpdpsl kqqsagnl 300

QY 334 FKLEYDGVFNKYTAQFRLNTKIGSRKIINRNYQFNGLSLNPYTNLNLTAAYNSGRQKY 393
D 301 fkleydgvfnkytaqf rldntkigsrkilnrnyqfn gylslnpytnlnltaaynsgrqky 360

QY 394 PGSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRETELTQTTLGFGNYFHFNEYKGRFPE 453
D 361 pgskf t g w g l l k d f e t y n n a k i l d l n n t a t f r l p r e t e l t q t t l g f g n y f h f n e y k g r f p e 420

QY 454 ELGLFFDPDQDNGLSYLGFRFKDGKGLLPQKSTIVOPAGSQYFNTFYFAALKKDIYRL 513
D 421 elglffdpdq dnglsylgrfkdgkgl lpqkstivopag sqy fntfyfdaal kkd i y r l 480

QY 514 NYSTNTVYRFGGEYTGYYGSDDEFFKRAFGENSPYKHKHNRSCGIYEPVLKYYKKRAN 573
D 481 nystntvy rfggey t g y y g s d d e f f k r a f g e n s p y k h k h n r s c g i y e p v l k y y k k r a n 540

QY 574 NHSVTSADFCDFMPFPASYSRTHRMPNIQEMFYFSQIGDSGVHTALKPERANTWQFGFNT 633
D 541 nhsvtsad f c d f m p p a s y s r t h r m p n i q e m f y f s q i g d s g v h t a l k p e r a n t w q f g f n t 600

QY 634 YKGLLKQDDTLGLKLVGYRSRIDNYIHNVYKWKWDLUNGDIPIPSWVSTGLAYTTOHRNFK 693
D 601 ykgl l k q d d t l g l k l v g y r s r i d n y i h n v y k w k w d l u n g d i p i p s w v s t g l a y t t o h r n f k 660

QY 694 DKVHKGFELNVDYGRFETNLSYAYOKSTOPTNFSDASESPNNASKEDQLKQYGLSR 753
D 661 dkvhkgfel nvd ygrf etnlsyayokst optn f s d a s e s p n n a s k e d q l k q y g l s r 720

QY 754 VSALPRDYGRLEVGRWLGNKLTGGAMRYFGKSIRATAEERYIDGTNGGNTSNFROLGK 813
D 721 vsalprdygr levgrw lgnkltggamryfgks irataeeryidgtnggntsnfrolgk 780

QY 814 RSKQDFTLARQPLIFDFYAAEPKKNLIFRAEVKNLFDRLYIDPLDAGNDAATQRYSS 873
D 781 rskq d f t l a r q p l i f d f y a a e p k k n l i f r a e v k n l f d r l y i d p l d a g n d a a t q r y s s 840

QY 874 FDPDKDEDVTCNADKTLCKNGKYGKTSKVLTNFARGRTFLITMSYKF 921
D 841 fdpdkdedv tcnadk t l c k n g k y g k t s k v l t n f a r g r t f l i t m s y k f 888

RESULT 6
AAV38939
ID AAY38939 standard; Protein; 922 AA.
XX
AC AAY38939;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae antigen encoded by ORF133.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
```

```
PR 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
XX
XX
XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
XX Claim 4; Page 479; 524pp; English.
XX
XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX reading frames (ORFs) AA211972-212358. The antigenic proteins,
XX their fragments, their nucleic acids and antibodies are used for
XX diagnosis, prevention (as vaccines) or treatment of Neisseria
XX infections, such as meningitis, septicaemia and gonorrhoea. Both
XX organisms are closely related. Fragments of the nucleic acids
XX are useful as hybridisation probes and antisense reagents.
XX
XX Sequence .922 AA;

Query Match 96.0%; Score 4696.5; DB 20; Length 922;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 883; Conservative 15; Mismatches 23; Indels 1; Gaps 1;

QY 1 MRSSFRLLKPTCYLMGVTLHYSAEDAGRAGSEAOQLVLEHVHAKRVPKOKKVTPTDA 60
Db 1 mrsfrllkpcfyimgvmlhyhsyaedagragseaoqlvledvhvakrpkdkkvtftda 60
QY 61 RAVSTRODIFKSSBNLNIIVRSIPGAFQODKSSGIVSLNIRGDSGFRVNTMVDGTTQT 120
Db 61 ravstrdqvfksgenlnivrsipgafqodkssgivslnirgdsgrvntmvdgittqt 120
QY 121 FYSTDAGRAGSSQFGASVDSNFIAGLDVVKGSGSAGINSLAGSNLRTLGVDVDDV 180
Db 121 fystdagragsqfgasvdsnfiaagldvvkgsfsgsaginslagsnrlrtlgvddvv 180
QY 181 OGNNYTGILLKGLTGNTSTGNMAAATGARKWLESAGSVGLYGHSSRVAQNYRVGGGG 240
Db 181 qgnntygillkgltgtustkgnmaaatgarkwlesagsvglvghssrgvaqnyrvggg 240
QY 241 QHIGNFAEYLERKQRYFQEGALKNSDSQKWERDLQKQWKPKYKNNN-OELQKY 299
Db 241 qhignfgeeylerkqryfgegalknagsgkwerdlqkqwkkyknnn-oelqky 300
QY 300 IEGHDKSWRENLAPOYDITPIDPSLLKQSQAGNLFKLEYDGVFNKNTAQFRDLNKTGSR 359
Db 301 ieehdkswrenlapqyditpidpsglkqqsagnlnlleydgvfnkntaqfrdlntrigr 360
QY 360 KIINRNYQFNVGLSLNPTNLNTAAYNSGROKYPKSGKETGWLKDKDFETYNNAKLTDL 419
Db 361 kiinrnyqfnvglslnptnlntaaynsgrkypkgakftgwgllkdfetynnakltldl 420
QY 420 NNTATFLPRETELQTLTGLGFNFHNEYKGNRFPPEELGLFFDQDQDNGLYSYLGRFGDK 479
Db 421 nntatflpretelqtltnfnyfhneygknrfpeelglffdgpdqdnlyslgrfkgdk 480
QY 480 GLLPQKSTIVOPAGSQVFNFTFYDAALKKDIYRLNLNTVTVYRFGGEYTYGYSDDDEFK 539
Db 481 glipqkstivopagsqvfnftfydaalkkdiyrlnlnstnainyrfggytygyseneff 540
QY 540 RATEGNSPTVKKHCRNSCGIYEPVLKYGKRRANNHVSISADPGDYFMPFASYSRTHRM 599
Db 541 ratgenspaykehdpcsgiyepvlkkygkrrannhsvsasadgdyfmpfagysrthrm 600
QY 600 PNQEMVFSQIGSGVHTALKPRANTWQGFNTYKGLLKQDDTLGLKLVGYRSRIDNY 659
Db 601 pniqemvfgsgvhtalkperantwqgfntykglkqddtlgklvgyrsridny 660
```

```
QY 660 IHNVYKWDNLNGDIPSWSVSTGLAYTIOHRNFKDKVHKHGFEELELNVDYGRFFTNLSYA 719
Db 661 ihnvgykwdnlngdipsvswstglaytirhnfkdkvkhkgfelelnvdygrfftnlsya 720
QY 720 YQKSTQPTNFSDASESNNASKEDQLKQGYGLSRVSALPRDYGRLEVTGTRWLGKLTIGG 779
Db 721 yqkstqptnfsdasespnnaskedqlkgyglsrvsalprdygrlevgtgrwlgknkltlgg 780
QY 780 AMRYFGKSIIRATAFERIYDCTNGNTSNFROLGKRSLIKQETLAROPLIFFDFAAYEPKK 839
Db 781 amryfgksirataeeryidctngntsnfrolgkrslikqetlarqpliffdfayayepkk 840
QY 840 NLIFRAEVKNLFORRYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNKGYYGGT 899
Db 841 nlifraevknlforryidplidagndaatqryysfdpkdkdedvtcnadkctlcnkgyggt 900
QY 900 SKSVLTNFAARGTFLTMSYKF 921
Db 901 sksvltnfargtrflmtmsykf 922

RESULT 7
AAY38938
ID AAY38938 standard; Protein: 871 AA.
XX
XX AC AAY38938;
XX
XX DT 08-OCT-1999 (first entry)
XX
XX DE N. meningitidis strain A antigen encoded by a partial ORF133.
XX
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX treatment; Neisseria infection; meningitis; septicaemia; gonorrhoea.
XX
XX OS Neisseria meningitidis.
XX
XX PN WO9924578-A2.
XX
XX PD 20-MAY-1999.
XX
XX PF 09-OCT-1998; 98WO-IB01665.
XX
XX PR 01-SEP-1998; 98GB-0019016.
XX PR 06-NOV-1997; 97GB-0023516.
XX PR 14-NOV-1997; 97GB-0024190.
XX PR 18-NOV-1997; 97GB-0024386.
XX PR 27-NOV-1997; 97GB-0025158.
XX PR 10-DEC-1997; 97GB-0026147.
XX PR 14-JAN-1998; 98GB-0000759.
XX
XX PA (CHIR-) CHIRON SPA.
XX
XX PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
XX WPI; 1999-327407/27.
XX PR N-PSDB; AA212353.
XX
XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
XX PS Claim 4; Page 477; 524pp; English.
XX
XX CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX reading frames (ORFs) AA211972-212358. The antigenic proteins,
XX their fragments, their nucleic acids and antibodies are used for
XX diagnosis, prevention (as vaccines) or treatment of Neisseria
XX infections, such as meningitis, septicaemia and gonorrhoea. Both
XX organisms are closely related. Fragments of the nucleic acids
XX are useful as hybridisation probes and antisense reagents.
XX
XX Sequence 871 AA;
```

```
Query Match      86.9%; Score 4348.5; DB 20; Length 871;
Best Local Similarity 94.6%; Pred. No. 0;
Matches 824; Conservative 8; Mismatches 38; Indels 1; Gaps 1;

QY 52 KDKKFTDARAVSTRODIPKSSNLDNIVSIPGAFQQDKSSGIVSLNIRGDSGFRGN 111
DB 1 kdkkfvdaravstrdqdfksxenldnivrpgafxqkssgxvslnirxdsfgfrvn 60

QY 112 TMVDGITQTFYSTDAGRAGSSQFGASVDSNFIAGLDVYVKGFSGAGINSLAGSANL 171
DB 61 tmvdgitxtfststdagragssqfgasvdsnfagldvkvgsfagsaginslagsanl 120

QY 172 RTLCGVDDVQGNNTYGLLLKGLTGTNSTKGNMAAAGARWLESGASVGVLYGHSRRSVA 231
DB 121 rtlxddvvdvgnxtgylgllkglgtgstkgnamaaigarkwlesgasvgvlyghsrrsva 180

QY 232 QNTRVGGGGQHIGNFGAEYLERKQRYFVQEGALKFNDSGKWERDLQROQWKYKPKNY 291
DB 181 qnrvvgggqhignfgaeylerkryfegqglkfnsngkwerdfqskwtkyqky 240

QY 292 N-NOELQKYIEGHDKSWRENLAPOYDITPIDPSLSKQOASAGNLFKLEYDGVFNKYTAQFR 350
DB 241 dapqelqyieghdkswrenlapqyditpidpslskxsaglnfkleydgvfnkytaqfr 300

QY 351 DLNATKIGSRKLIINRNQYFNGLSLNPYTNLNTAAVNSGRQKYPKSGKFTGWGLLDKFET 410
DB 301 dlnatkigsrkliinrnqyfnlslnpynlnlntaaavnsgrqkypksgkftgwlxkdfet 360

QY 411 YNNAKILDLNATFRLPRETEQLTTLGFNYFHNEYGNRFPPEELGLFFQDPDODNGLYS 470
DB 361 ynnakiildlntxtftrpreteqltltgfnfneygknrfpeelglffqdpdxndnglys 420

QY 471 YLGRFKDGKLLPKQKSTIVOPAGSYQNTFYFDAAKKDIYRLNYSNTVYGRFGGEYTG 530
DB 421 ylgrfkdgkllpqkstivpagsyqntfyfdaalkkdiylrlnysntvgyrfggxytg 480

QY 531 YGSDDEFKFAFGENSPTYKKHCRSGIYEPVLKKYGGKRRANNSHVSISADFGDPMFP 590
DB 481 ygsddefkfafgensptykxhcnsgciyepvlkkygkrrannshvsisadfgdymfp 540

QY 591 ASYSRTHRMFNIQEMVFSQIGSGVHTALKPERANTWQGFNTYKKGLLKODDTLGLKL 650
DB 541 asysrthrmfniqemvysqigsgvhtalkperantwqgfntykykgllkqddilglklv 600

QY 651 GYRSRIDNYIHNVYKWDMLNGDIPSWVSSTGLAYTTIQHRNFKDKVHKHGFLELNYDYG 710
DB 601 gyrsridxyihnvyykwdmlngnipswvsstglaytiqhrnfkdvkhkhgflelnydyx 660

QY 711 RFTNLSYAYOKSTOPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWR 770
DB 661 rftnlsyayokstoptnfsdasespnnaskedqlkqgyglslrvsalsprdygrlevgtwr 720

QY 771 LGNKLTLGGAMRYFGKISIRATAERYIDGTNGGNTSNFROLGKRISIKOTETTLAROLIFD 830
DB 721 lgnkltlggamryfgkisirataeryidxtngxtnsnfrqlgkrsixqetclarqlifd 780

QY 831 FYAAEPKKNLIFRAEYKKNLFDRRYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKT 890
DB 781 xyaaepkknlifraeykknlfdrriyidpldagndaatqryyssfdpkdkdeevtcndnt 840

QY 891 LCNGKYGGTGSKSLVTLNPARGRTFLITMSYKF 921
DB 841 lcngkyggtgsksltnfargxtflitmsyxf 871
```

```
RESULT      8
AAY44428
ID AAY444428 standard; Protein: 947 AA.
XX
AC AAY444428;
XX
DT 22-MAR-2000 (first entry)
```

```
XX M. catarrhalis (ATCC 43617) BASB021 polypeptide.
DE BASB021; HasR; outer membrane haem-binding protein; sinusitis;
XX otitis media; pneumonia; nosocomial infection; auditive nerve damage;
KW delayed speech learning.
KW Moraxella catarrhalis.
OS
XX
XX WO9964602-A2.
XX
XX 16-DEC-1999.
XX
XX 31-MAY-1999; 99WO-EP03824.
XX
XX 09-JUN-1998; 98GB-0012440.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Thonnard J;
XX
XX WPI; 2000-116545/10.
DR N-PSDB; AAZ29682.
XX
XX New isolated Moraxella catarrhalis BASB021 polynucleotides, used to
PT develop products for the diagnosis, prevention and treatment of
PT infections causing e.g. otitis media -
XX
XX Claim 1; Page 81-84; 87pp; English.
XX
XX The present sequence is BASB021 polypeptide, which shows sequence
CC homology with Serratia marcescens HasR outer membrane haem-binding
CC protein. It is encoded by DNA from Moraxella catarrhalis strain Mc2931
CC (ATCC 43617). BASB021 polynucleotides and polypeptides may be used for
CC prognosis, staging of disease, determining response to drug treatment
CC and drug screening. They can be used in vaccines for generating an immune
CC response and for preventing or treating Moraxella infections which may
CC cause otitis media, pneumonia, sinusitis, nosocomial infections and
CC invasive diseases, auditive nerve damage, delayed speech learning,
CC infection of upper respiratory tract and inflammation of the middle ear.
CC Anti-BASB021 antibodies can be used to diagnose and treat Moraxella
XX infections.
XX
XX Sequence 947 AA;
SQ

Query Match      52.9%; Score 2588.5; DB 21; Length 947;
Best Local Similarity 55.6%; Pred. No. 7.4e-179;
Matches 503; Conservative 144; Mismatches 232; Indels 25; Gaps 12;

QY 39 VLEDVHVKARR-VPKDKKVTFDARAVSTRODIPKSSNLDNIVSIPGAFQQDKSSGIV 97
DB 48 ildevvvtatngtkksqpkfktasatsvrenvfnaenidaivrsvpvgaftqgdkssglv 107

QY 98 SLNIRGDSGFRVNTVMYDGTOTFYSTSDAGRAGSSQFGASVDSNFIAGLDVYVKGFS 157
DB 108 slnrvgsdgrfnsmdvgtvtqtfyststdagrsggtsqfgavldqfagvelnksfn 167

QY 158 GSAGINSLAGSANLRTLCGVDDVQGNNTYGLLLKGLTGTNSTKGNMAAAGARWLESGA 217
DB 168 gkgglntltsanfrtlnaddvikdknfgiakglgkgnatdknfnlaagrgwldng- 226

QY 218 SVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERKQRYFVQEGALKFNDSGKWERD 277
DB 227 sisalyayshkdisqnykvvgggthignvgddlllskqkvfakehaltynearsrqkd 286

QY 278 LQROQ-----W--KYK-PYKNY-----NNQELQKYIEGHDKSWRENLAPOYDITPID 321
DB 287 ltkldketgkplwdrkyvgfggkcyglgcidtkekfdeyvadvkqgqwkghakeysitpid 346

QY 322 PSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNATKIGSRKLIINRNQYFNGLSLNPYTNL 381
DB 347 italnqtkskshlakirynndtsdvglqlrkmtdttigrrisrindndyqidaaynpneiidlk 406
```


Db 565 epilhsghkkaafnhsatlsaelisdyfmpfftyrthrmpnigemffsfqvsngvntalk 624
 QY 621 PERANTWQFENYKKGKLLKQDDTLGLKLVGYSRIDNYLHNVYGGKWDNLNGDIPSWSS 680
 Db 625 pedsdyqlgintykgllftqddvlgklygysfknynhnygvww--rdgmptwaes 682
 QY 681 TGLAYTIOHRNFKDKVHKHGFELNLDYGRFTNLISYAYOKSTQPTNFSDAESPNNAS 740
 Db 683 ngfkytiiahnykpvkksveleindymgrffanvsyayqrtnqpcnycadasprnnas 742
 QY 741 KEDOLKOGYGLSRVSLPRDYGRLEVRWLGNKLTGLGAMRYFGKSRIRATAERYIDGT 800
 Db 743 qedilkgyglsvmlpkdygrlelgrwfdqkltglalaaryygsksraticeeyings 802
 QY 801 NGGNTSFROLGRKRSIKQTETLARQPLIFDFYAAYPEKKNLIFRAEVKNLFDRIYDPLD 860
 Db 803 r-fkntllrrenyavvktedikkqplildhvsyepiklilikaevgnlldkryvdpld 861
 QY 861 AGNDAATORYSSFDPKDDEYTCNADKTLGNGKYGTSKSVLTNFARGRTFLITMSYK 920
 Db 862 agndaasqryssl-----nnsiecaqdsac----ggsdktvlynfargrtvylisinyk 912
 QY 921 F 921
 Db 913 f 913
 RESULT 10
 AAY94672
 ID AAY94672 standard; Protein; 918 AA.
 XX
 AC AAY94672;
 XX
 DT 01-DEC-2000 (first entry)
 XX
 DE Haemophilus antigen BASB070 protein sequence.
 XX
 KW Vaccine; BASB070; Haemophilus influenzae; strain nH3224; pneumonia;
 KW chronic bronchitis; sinusitis; otitis media; meningitis; antigen;
 KW systemic disease; outer membrane protein.
 XX
 OS Haemophilus influenzae.
 XX
 PN WC200050599-A1.
 XX
 PD 31-AUG-2000.
 XX
 PF 22-FEB-2000; 2000WO-EP01423.
 XX
 PR 24-FEB-1999; 99GB-0004183.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Ruelle J, Thonnard J;
 XX
 DR WPI; 2000-543912/49.
 DR N-PSDB; AAA28031.
 XX
 PT Vaccinating against Haemophilus influenzae using BASB070 polypeptides
 XX and/or the nucleic acids that encode them -
 PS Claim 1; Fig 2; 97pp; English.
 XX
 CC This invention relates to a vaccine composition which contains BASB070
 CC polypeptides from Haemophilus influenzae and/or the nucleotide sequences
 CC that encode them. BASB070 DNA sequences encode proteins which are
 CC integral outer membrane proteins with a beta-barrel conformation. The
 CC invention includes BASB070 nucleotide and protein sequences from
 CC H. influenzae strains RdKw20 and nH3224. The invention includes an
 CC expression vector and recombinant live microorganism comprising a BASB070
 CC polynucleotide sequence, a process for the production of a BASB070
 CC protein, an antibody specific for the two BASB070 proteins of the
 CC invention, and a method for diagnosing an H. influenzae infection.

CC Haemophilus influenzae is a common cause of pneumonia, exacerbation of
 CC chronic bronchitis, sinusitis and otitis media, and H. influenzae type b
 CC causes bacterial meningitis and systemic diseases. The vaccine exhibits
 CC antibacterial activity, and is used in the preparation of an agent for
 CC use in generating an immunological response in a mammal.
 CC The present sequence represents a BASB070 protein isolated from
 CC H. influenzae strain nH3224. The protein is used in the production of
 CC the vaccine of the invention.
 XX
 XX Sequence 918 AA;
 QY
 Query Match 50.7%; Score 2483.5; DB 21; Length 918;
 Best Local Similarity 51.1%; Pred. No. 2.9e-171;
 Matches 494; Conservative 154; Mismatches 225; Indels 93; Gaps 17;
 QY 1 MRSFRLKPCIFYL---MGVTLHYHYAEADAGRAGSQAQIQVLEDVHVAKARVPDKKVF 57
 Db 1 mkaikaiknltlslintigmtitqaqaetlg-----qidvvekv-----isndkkpf 48
 QY 58 TDARAVTQDIFKSENLENLIVRSIFCAFTQODKSSGIVSLNIRGDSGFGVRVMTWDGI 117
 Db 49 teakakstrenvketqtdidqvirspgafqgdkgsvsvnirgenglgrvntmvdgv 108
 QY 118 TQTFYSTDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVD 177
 Db 109 tqtfystaldsgsgsgsgfgaaidpniagvdknfnsgasinalagsanlrltvsn 168
 QY 178 DVQGNNTYGLLLKGLTGTNSTKGNMAAAGARKWLSGASGVLYGHRSRVSQAQNYRVG 237
 Db 169 dvitddkpfgiikgmtgsnatksnfmttaagrkwldnggyvgyvgsqrevsqdyri- 227
 QY 238 GGGQHGICFCAEYLERRKQRYFVQEGALKFNSDGKWERDLQROOW----- 283
 Db 228 gggerlaslsgqdlakekeikfrndgyv-lns-agqwapdlnkphwscntpslkdksms 285
 QY 284 -KYPYK-----NYYNQELQY---IEGHDKSWRENLAPOYDIT 318
 Db 286 tsckpyrlgpaatrqellkelledgkpkdieklqkndgieteeksfen-kdydva 344
 QY 319 PIDPSSILKQOSAGNLKFLKLEYDGVFNKYTAQPRDLNTKIGSRKIINRYQFNYGSLNPT 378
 Db 345 piepgslqsrshrshllkfeysddhhtlgaqirtldnkigsrkienryqvynfnnsyl 404
 QY 379 NLNLTAAYNGRQYKPGSKFTGWLKDKDFETYNNAKLLDNTATFRLPRETELQTTLG 438
 Db 405 dlnlmaahngktlypkggffagvgvadklltknvanilvinnshntflilkeidlktllg 464
 QY 439 FNYFHEVYGNRPPEELGLFEDGPDQDNGLYSY--LGRFKGDKGLLPKSTIVOPAGSOY 496
 Db 465 fnyftneyksnrpfpeelsilfyneshdqgylslsnkgrysgskglpqrsvillqpsgkqk 524
 QY 497 FNTFYDAALKKDIYRLNYSYNTVTGYRFGGTYGYYSDDDEFKRAFGSENSPTYKHKCNRS 556
 Db 525 fktvyfdalskgyihlnysvntfhyafngyvykntadk----- 565
 QY 557 CGIYEPVLKXYGKKRANNHVSISADFGYFMPFASYSRTHRMNIOBMYSQJGDSGVH 616
 Db 566 --lneplhskghkafnhsatlsaelisdyfmpfftyrthrmpnigemffsfqvsdavn 623
 QY 617 TALKPERANTWQFENYKKGKLLKQDDTLGLKLVGYSRIDNYLHNVYGGKWDNLNGDIPS 676
 Db 624 talkpedsdyqlgintykgllftqddvlgklygysfknynhnygvw--wsrdgvmpe 682
 QY 677 WVSSTGLAYTIOHRNFKDKVHKHGFELNLDYGRFTNLISYAYOKSTQPTNFSDAESP 736
 Db 683 warlmgfrltiahnykpvkksveleindymgrffanvsyayqrtnqpcnycadasarp 742
 QY 737 NNASKEDQLKQGYGLSRVSLPRDYGRLEVRWLGNKLTGLGAMRYFGKSRIRATAERY 796
 Db 743 rnaskeelkqgyglsvmlpkdygrlelgrwfdqkltglalaaryygsksratteqee 802
 QY 797 IDGTN-CGNTSNFROLGRKRSIKQTETLARQPLIFDFYAAYPEKKNLIFRAEVKNLFDRIY 855


```

      803  ingryeknttrdriy--yaikkteekkpilidllhvsyepikdliikaevqalldkry 860
      856  IDPLDAGNDAATORYYSSFDPKDDEVTNADKTLGNGKYGGTSKSVLTNFAARGTFLI 915
      861  vdpldagndaasqrrysl-----ndslackinestcn---dsektvlynfargrtyil 912
      916  TMSYKF 921
      913  slnykf 918

RESULT 11
AA38936
ID AAY38936 standard; Protein; 393 AA.
XX
AC AAY38936;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis antigen encoded by a partial ORF133.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria meningitidis.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WPI; 1999-327407/27.
DR N-PSDB; AAZ12351.
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
PS Claim 4; Page 473; 524pp; English.
XX
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AAZ11972-Y12358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 393 AA;

Query Match 40.8%; Score 1996; DB 20; Length 393;
Best Local Similarity 95.2%; Pred. No. 1.9e-136;
Matches 373; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

QY 530 GYGSDDDFKRAFGNSPTVKKHNCRCGIYEPVLKKYKKRANNSVSIASDFGDFMP 589
      |||||||
      2 gyygsdddfkrafgensptvkkhncrcgiyepvlkkykkrannhsvsisadfgdfmp 61

```

```

QY 590 FASYSTRHRMPNIQEMYFSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKL 649
      |||||||
      62 fasystrhrmpniqemyfsqigdsgvhtalkperantwqgfntykgkllkqddtlglkl 121
QY 650 VGRSRDNTYHNHYGKWDNLGNDIPSWSVSTGLAYTIOHRNFKDKVHKHGFLELNDY 709
      |||||||
      122 vgrsrldntyhnvygkwdnlngndipswsvstglaytihrxfkdkvhqkxxxxxydy 181
QY 710 GRPFTNLSYAVOKSTOPTNFSDASESPNNASKEDQLQOGYGLSRVSALPRDYGRLEVGR 769
      |||||||
      182 grpftnlsyavokstoptnfdsasespnnaskedqlqgygysrvsalprdygrlevgr 241
QY 770 WLGNKLTGGMARYFGKSIRATAEERYIDGTNGGNTSNFQLGKRSIKQTETLARQPLIF 829
      |||||||
      242 wlgnkltlgamryfgksirataeeryidgtnggntsnfqlgkrsikqetlarqplix 301
QY 830 DFYAAEPEKKNLIFRAEVKNLFDRIYDPLDAGNDAATORYYSSFDPKDDEVTCTNADK 889
      |||||||
      302 dfyaaepkknlfraevknlfdrdyidpldagndaatoryyssfdbpkdkdxvctcnadk 361
QY 890 TLCNGKYGGTSKSVLTNFAARGTFLITMSYKF 921
      |||||||
      362 tlcngkyggtksvltnfargrtflitmsykf 393
DB
RESULT 12
ABB52934
ID ABB52934 standard; Protein; 753 AA.
XX
AC ABB52934;
XX
DT 11-FEB-2002 (first entry)
XX
DE Escherichia coli polypeptide SEQ ID NO 1277.
XX
KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicaemia;
KW pyelonephritis; antibiotic resistance.
XX
OS Escherichia coli.
XX
PN WO20016572-A2.
XX
PD 13-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-EP03445.
XX
PR 10-MAR-2000; 2000PR-0003145.
PR 02-FEB-2001; 2001PR-0001449.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
DR WPI; 2001-550253/61.
XX
PT A library of DNA fragments of Escherichia coli strains for the
PT phylogenic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -
XX
PS Example 6; Fig 6; 646pp; English.
XX
CC The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC and encoded proteins (ABB52459-ABB52919) of nature B2/D+A-. The
CC polynucleotides have potential antiinflammatory, antibacterial and
CC immunosuppressive activity as part of pharmaceutical compositions used to
CC treat, palliate or prevent extra-intestinal E. coli infections. The
CC polypeptides are useful for determining the phylogenic group of a given
CC E. coli strain. These polypeptides can detect and treat an undesired
CC development of E. coli, particularly an extra-intestinal infection that
CC include systemic and non-diarrhoeal infections such as septicaemia,

```


CC pyelonephritis and meningitis this is particularly advantageous as
CC bacterial resistance is increasing with the more frequent use of broad
CC spectrum antibiotics.
XX
SQ Sequence 753 AA;

Query Match	18.3%	Score 998;	DB 22;	Length 753;
Best Local Similarity	28.7%;	Pred. No. 1.9e-56;		
Matches 262;	Conservative 118;	Mismatches 296;	Indels 236;	Gaps 26;

Qy	45	VKAKRVPDKKVVTDRAVSTRQDIPKSSSENLDNIVRSPGCAFTQODKSSGIVSLNIRGD	104
Db	43	vsvgtktseaelektgatsr-ttdknlqslatvrsmpgtqtiqdpqggasvnaigrm	101
Qy	105	SQFGRVMTWGDITQTFYSTSDAGRAGGS--SOFGASVDSNFIAGLDDVVVKSGFSGSAGI	162
Db	102	sgfrvrtmvdgtiqtsfygtsstgthstnmmagvlidpnlllavdvtrgdsgsesi	161
Qy	163	NSLAGSANLRTLGVDDVVVOGNNTYGLLLKGLTGTSTKGNAMAAIGARK--WLEGSASYG	220
Db	162	nalagsanmtigtvddvifngnfygrsrfsvgsnglrgsmialggksdafdtgtg-slg	220
Qy	221	VLYGHSRRSAQNYRVYCGGQGHGICNGAEVLERRKQRYFVQEGALKFNSDSGKWERDLQ	280
Db	221	vmaavsgssvysnfsgsg-----	239
Qy	281	QOWKYKPYKNYNQELQYIEGHDKSWRENLAPOYDITPIDPSSLXQOAGSLFKLEY-D	339
Db	240	-----inskef-----gydk-----ymkqnpksqlykmdirp	266
Qy	340	GVFNKYTAQFRDLNWKIGSKRIINRYQFNWGLSLNPNYNL---NLTAAYNSGRQKYPKG	396
Db	267	defnsfalsartyenkfrdiitsdddyiky--hytfselidfnvtastsgnqkyrdg	324
Qy	397	SKFTGWLKDDFETYNNAKLTNNATPRLPRETELQTTLGPNYFHNVEYGNRRPEELG	456
Db	325	slyfyf---ktaqnsrdaldmnrstfv-adndlefmlgsklmrty-----	369
Qy	457	LFFDGPDDNGLYSYLGRFGDKGLIPQKSTIVOPAGSYFNTFYFDAALKKDI---YR	512
Db	370	-----drtihsaagdpkqanesi---ennpfapsggqdisaalytqlkvtrgiweadfn	419
Qy	513	LNYSNTNV-GYRFGGEY-----TGYGSDDEFKRAFGENSPYKHCNRCGIEYEPVLK	565
Db	420	lnytrnitykpacdservicvpqsgsyldld-----	450
Qy	566	KYGKRRANNHSVISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHHTALKPERAN	625
Db	451	---kegafnpsvqisaqvtpwlpqfigyskrmaphniemffnsngsgasmnplfkperae	507
Qy	626	TWQGFNTYKGLLKQDDTTLGLKLVGVRSDINVIHN-----	699
Db	508	twqagfnldrllveqdalrfkalayrsirqnyiyysesylvcsgrkcsklspevigngwe	567
Qy	670	LNQDIPSWSSSTGLAY*IQHRNFKDKVHKHCFEELNLDYDGRFETNLVAYOKSTOPTNF	729
Db	568	---gisdeysdmnyiv-----nsasdrviakfgelemdydagfagrlsfsqqtdqpts	620
Qy	730	SDASESPNNAKEDQLKQGYGLRVSALPRDYGRLEVGTRWLGNKLTLLGGAMRYFGKSTIR	789
Db	621	asth-----fgagditelprkymtldgtvfrffdnaltltiikytgkarr	665
Qy	790	ATAERYIDGTNGNTSNFQRLGKRSIKQETLAROPLDFDYAAEYPPKNLIFRAEVKN	849
Db	666	lspfedqentga-----likg--dlpqtiatliidtygyeyfrnltklisvgn	711
Qy	850	LFDRRYIDPLDAGNDAATQRYYSFDPKDKDEVDVTCNADKTLNCGYKGYGTSKSVLTNFA	909
Db	712	lmnrdyseainkln-----mmpqigdehthpans-----ar	741
Qy	910	GRTELIITMSYKF	921

Db 742 grtwifggdirf 753

RESULT 13
ABB52968
ID ABB52968 standard; Protein; 753 AA.

AC	ABB52968;
XX	
DT	11-FEB-2002 (first entry)
XX	
DE	Escherichia coli polypeptide SEQ ID NO 1277.

DE Escherichia coli polypeptide SEQ ID NO 1277.

Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 systemic infection; non-diarrhoeal infection; septicaemia;
 pyelonephritis; antibiotic resistance.

OS *Escherichia coli*.

PN WO200166572-A2.

PD 13-SEP-2001.

PF 12-MAR-2001; 2001WO-EP03445.

PR 10-MAR-2000; 2000FR-0003145.

XX
XX

[illegible]XX
XXXX
XX

PT phylogenetic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -

PS Example 6; Fig 6; 646pp; English.

The invention relates to a library of DNA fragments of *Escherichia coli* strains comprising polynucleotides (ABR89577-ABR88729 and ABR89533) and encoded proteins (ABBS2459-ABBS22919 and ABBS2954-ABBS3094) of nature B2/DvA-. The polynucleotides have potential antinflammatory, antibacterial and immunosuppressive activity as part of pharmaceutical compositions used to treat, palliate or prevent extra-intestinal *E. coli* infections. The polypeptides are useful for determining the phylogenetic group of a given *E. coli* strain. These polypeptides can detect and treat an undesired development of *E. coli*, particularly an extra-intestinal infection that include systemic and non-diarrhoeal infections such as septicæmia, pyelonephritis and meningitis this is particularly advantageous as bacterial resistance is increasing with the more frequent use of broad spectrum antibiotics.

AA	Sequence	753 AA;
SQ		

Query Match	18.3%	Score 898;	DB 22;
Best Local Similarity	28.7%	Pred. NO. 1.9e-56;	
Matches 262;	Conservative 118;	Mismatches 296;	
Indels 236;	Gaps 26;		

Qy 45 VKAKRVPKDKKVFTDARAVSTRQDIKFSSSENLDNIVSRIPGAFQTQODKSSGVISLNRGD 104
| : :: | | : : : | | : : : | : : : | : : : |
Db 43 vsvgkttsecealektgatsr-ttdknqlsqidatvrsmpgytqidpgggaisvnrigm 101

Qy	105 SGGRVNTMVDGITOTFYSTSDAGRAGGS--SQFGASVDSNFIAGLDVVVKGSFSSSAGI 162
Db	102 sgfrvntmvdgitqsfygtstgthtghstnmgavlidpnllivadvtrgdsssgsegI 161

Qy 163 NSLAGSANLRLTGLVDVVQGNNYTGLLKGLTGTNSTKGNAWAIGARK--WLESGASVG 220
:
Db 162 nalagsanmrftigvddvifngntqyrlrsrfsvsgnqlgrsmialgoksdafdtgt-sig 220

Qy	221	VLYGHSRRSVAQNRYVGGCGQHICNGAELERPKORYFVQEGALKFNSDSGKWERDLQR	280
Dd	221	vmaavsgssvysnfsngsg	239
Qy	281	QOMWYKPYKNNOELQKYIEGHDKSWRENLAPOYDITPIDSSLKQQOSAGNLFKLEY-D	339
Dd	240	-----inskef-----gydk-----ymkqnpksqlykmdirp	266
Qy	340	GVENKYTAQFRDLNTKIGSRKIINRNYOFNYGLSLNPYNL---NLTAAYNSGRQKYPKG	396
Dd	267	defnselsartyenkfrirditsdyiky--hytpfselidfnvtastsgnkdyrdg	324
Qy	397	SKFTGWLLKDFETNNNAKTILDJNDTATRLPRELOTTLGTFNVPHNBYGNKRPEELG	456
Dd	325	slytfy---ktsaqnrdsaldinntsrftv-adndlefmjgsklmrty-----	369
Qy	457	LFFDPDODNGLSYLGRFKDGKLLPQKSTIVOPAGSQFYNTFFDAALKXDI----	512
Dd	370	-----drtihsaagdpkandesi---enmpfapsggqdialyglkvtrgiweadfn	419
Qy	513	LNKSTNTV-GYRFGEY-----TGYYGSDEPFKRAFGENSPTYKKHCNRSCGIYEPVLK	565
Dd	420	lntytrnitgypacodrsrvicvpqgsydd-----	450
Qy	566	KYGKRRANHSVSIADFGDYFMPPFASYSRTHRMPIQEMIFYSOIGDCSVHTALKEPERAN	625
Dd	451	--keggfnpsvqlsaqtwlpqpfifygksmrapiqemifsnsggasmnpflkperae	507
Qy	626	TWQGFPNYKXGLLKDDTLGLKLGVYRSRIDNYIHN-----VYKGWMD	669
Dd	508	twdagfnldtrdlvegdalrfkalayrsriqnuyisesylvycsggrkcslpevlngwge	567
Qy	670	LNGDIFSWWSSTGLATIQHNFKDVKVHKHGELELNVDYGRFFTNLSVAYOKSQTPTNF	729
Dd	568	--gisdeyadmnyiyv---nsasvdiakfelemdydagfafgrlsfsgqdtqpts	620
Qy	730	SDASESPNNASKEDQLKQGVLGSRVSALPDYGRLEVGTRELGNKLTJLGAMRYPGKSIR	789
Dd	621	asth-----fgaedtelcpkytmtdtgvrfdnaltgtiiiklycgkarf	665
Qy	790	ATAEERYIDGTNGCNTSNFRQLGKRSIKQETLARQPLIFFDYAAYPEKKNLIFRAEVKN	849
Dd	666	lspdfegdehtga-----likq-dlpqiptiidlygteyeynrnltklsvqn	711
Qy	850	LFDPRYIDPLDAGNDAATORYSSFPDKDDEBVTCTNADKTLCNCKYGGTSKVLTNFR	909
Dd	712	lmurdysealnkn-----mmmpjlgdethpans-----ar	741
Qy	910	GRTEFLITMSYKF	921
Dd	742	grtwifeggdirf	753

RESULT 14

RESOL 14
AAY07476
ID AAY07476 standard; Protein; 915 AA.

XX
AC AAY07476:

AC	AA10/470;
XX	
DT	17-AUG-1999 (first entry)

DE N.gonorrhoeae transferrin binding protein.

Transferrin binding protein; iron; outer membrane protein; uptake;
 Neisseria gonorrhoeae; nutrient; growth; Neisseria meningitidis; pathogen;
 receptor; antibody.

AA
OS

XX PN US5912336-A.

XX
PD 15-JUN-1999.

XX	23-DEC-1994;	94US-0363124.
PF	XX	
XX	23-DEC-1994;	94US-0363124.
PR	XX	
PR	23-AUG-1990;	90US-0572187.
PR	23-AUG-1990;	90US-0572187.
PR	05-NOV-1992;	92US-0973336.
PR	20-SEP-1993;	93US-0124254.
XX	XX	
XX	(UYNC-) UNIV NORTH CAROLINA.	
PA	XX	
XX	Cornelissen CN, Sparling PF;	
PI	XX	
XX	WPT; 1999-357219/30.	
DR	N-PSDB; AAX78929.	
DR	XX	
XX	Nucleic acid molecules encoding transferrin binding proteins	
XX	Disclosure; Fig 1; 38pp; English.	
PT	XX	
PS	XX	

Query Match 5.6%; Score 273.5; DB 20; Length 915;

Query Match 5.68; Score 273.5; DB 2

Best Local Similarity 21.3%; Pred. No. 4.6e-11;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRLKPICFYLMGVTLYHYSYAEDAGRAGSEAIQVLEDVHVKA -- RVPKDKKVFTDARA 62

```

Db
7 frlnlclslmtal---payaenv-qag-qaqekqldtiqvkakkqktrrdnevtglgkl 61

```

QY 63 VSTRQDIFKSSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGRVNTMVDGIT 118

Db 62 vktadtlskeqvldirdltrydpqiavveqqrqassq---ysirq-mdknrvsltvdqqla 117

QY 119 Q-TFYSTDAG--RAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTL 175

Db 118 qiqsytaqaalqqtrtaqssgaineiyevenvkaveiskqsnsv eqqsqalagsvafqtkt 177

QY 176 VDDVVQGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRSVAQNYR 235

```
Db      178 advdgearqwgq|-----qsktayvsknqr|-----tqslalagr 212
```

QY 236 VGGG-----GQHIGNFGAEYLERRRKORYFVOEGALKFN-----SDSGKWERDLQRQOW 283

Qy	1	MRSRFLKPICFYLMGVTLHYHYSAEDAGRAGSBAQIQVLEDVHVHKARVPKDKKVF7DA	60
Db	1	MRSRFLKPICFYLMGVTLHYHYSAEDAGRAGSBAQIQVLEDVHVHKARVPKDKKVF7DA	60
Qy	61	RAVSTRQDIFKSSENLDNIVRSIPCAFTQDDKSSGIYSLNIRGDSGFGRVMTWDGIIQT	120
Db	61	RAVSTRQDIFKSSENLDNIVRSIPCAFTQDDKSSGIYSLNIRGDSGFGRVMTWDGIIQT	120
Qy	121	FYSTSTADGRAGGSSQFGASVDSNFIAGLDVVKVGSFGSSAGINSLAGSANLRTLCGVDDV	180
Db	121	FYSTSTADGRAGGSSQFGASVDSNFIAGLDVVKVGSFGSSAGINSLAGSANLRTLCGVDDV	180
Qy	181	QGNNTYGLLLKGLTGTTSTKGNAMAAICARKWLESAGSVGLYGHRSRSVAQNTYRVGGG	240
Db	181	QGNNTYGLLLKGLTGTTSTKGNAMAAICARKWLESAGSVGLYGHRSRSVAQNTYRVGGG	240
Qy	241	QHINFGAEYLERRKQRYFVQEGALKFNDSGKWERDLQROQWKYKPYKNTYNNQELQYI	300
Db	241	QHINFGAEYLERRKQRYFVQEGALKFNDSGKWERDLQROQWKYKPYKNTYNNQELQYI	300
Qy	301	EGHDKSWRENLAPOYDITPDPSILKQOSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRK	360
Db	301	EEHDKSWRENLAPOYDITPDPSILKQOSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRK	360
Qy	361	IINRYOFNYGLSLNPVTNLNLTAAYNSGRGOKYKPGSKFTCGWLKKDLPFETVYNNAKILDLN	420

Db 361 IINRNYQFNYGLSNPYTNLNLTAAYNSGRQYKPKSGKFTGWGLLKDFETYNNAKILDNL 420
QY 421 NTATFRLPRETELOTTGLGFNYFHNEYKGNRPPEELGLFFDGPDOODNGLYSLGRFKDGK 480
Db 421 NTATFRLPRETELOTTGLGFNYFHNEYKGNRPPEELGLFFDGPDOODNGLYSLGRFKDGK 480
QY 481 LLPQKSTIVPAGSQYFNTFYDAALKKDIYRLNYSNTVYRGFGGYTGYGSDDEFKR 540
Db 481 LLPQKSTIVPAGSQYFNTFYDAALKKDIYRLNYSNTVYRGFGGYTGYGSDDEFKR 540
QY 541 AFGNSPTYKKHCNRSQGIYEPVLKYYGKKRANNSHSVISADFGDYFMPFASYSRTHRM 600
Db 541 AFGNSPTYKKHCNRSQGIYEPVLKYYGKKRANNSHSVISADFGDYFMPFASYSRTHRM 600
QY 601 NIQEMYSQIGDSGVHVALPERANTWQFGNTYKGLLKODDPLGLKLVYRSRIDNYI 660
Db 601 NIQEMYSQIGDSGVHVALPERANTWQFGNTYKGLLKODDPLGLKLVYRSRIDNYI 660
QY 661 HNVYCKWMDLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFELBNLDYGRFFTNLSYAY 720
Db 661 HNVYCKWMDLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFELBNLDYGRFFTNLSYAY 720
QY 721 QKSTQPTNFSASPNNAKEDOLKQGYGLSRVSALPRDYGRLEVTGRLGNKLTLLGGA 780
Db 721 QKSTQPTNFSASPNNAKEDOLKQGYGLSRVSALPRDYGRLEVTGRLGNKLTLLGGA 780
QY 781 MRYFGKSIRATAEERYIDGTNGNTSNFROLGKRSIKQETTLARQPLIFDFYAAAYEPKN 840
Db 781 MRYFGKSIRATAEERYIDGTNGNTSNFROLGKRSIKQETTLARQPLIFDFYAAAYEPKN 840
QY 841 LIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLCKNGKYGTS 900
Db 841 LIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLCKNGKYGTS 900
QY 901 KSVLTNFARGRTFLMTMSYKF 921
Db 901 KSVLTNFARGRTFLMTMSYKF 921

RESULT 2
G81865
Probable outer membrane substrate binding protein NMA1700 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: G81865
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
A:Reference number: A81775; MUID:20222556
A:Accession: G81865
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-922 <PAR>
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84928.1; PID:g738034
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: NMA1700

Query Match 97.3%; Score 4763.5; DB 2; Length 922;
Best Local Similarity 97.5%; Pred. No. 1e-290;
Matches 899; Conservative 9; Mismatches 13; Indels 1; Gaps 1;

QY 1 MRSSFLRKPCFYLMGVTLTHYSAEDAGRAGSEAOQVLEDVHVAKRVPKDKKVFDTA 60
Db 1 MRSSFLRKPCFYLMGVTLTHYSAEDAGRAGSEAOQVLEDVHVAKRVPKDKKVFDTA 60
QY 61 RAVSTRODIFKSSNLDNIYRSIPGAFQODKSSGIVSLNIRGDSGFGRTVWVDGTTOT 120
Db 61 RAVSTRODIFKSSNLDNIYRSIPGAFQODKSSGIVSLNIRGDSGFGRTVWVDGTTOT 120

QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLGVDVV 180
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLGVDVV 180
QY 181 QGNNTYGLLLKGLTGTNTSGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGGG 240
Db 181 QGNNTYGLLLKGLTGTNTSGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGGG 240
QY 241 QHIGNTGAELYERKKORYFVOEGALKFNSDGKWERDLQROOMKYKPKNYN-NOELQKY 299
Db 241 QHIGNTGAELYERKKORYFVOEGALKFNSDGKWERDLQROOMKYKPKNYN-NOELQKY 300
QY 300 IEHDKSWRENLAPOQYDITPIDPSSLKQKQAGNLFKLEYDGVFNKYTAQFRDLNTKIGSR 359
Db 301 IEHDKSWRENLAPOQYDITPIDPSSLKQKQAGNLFKLEYDGVFNKYTAQFRDLNTKIGSR 360
QY 360 KIINRNYQFNYGLSNPYTNLNLTAAYNSGRQYKPKSGKFTGWGLLKDFETYNNAKILD 419
Db 361 KIINRNYQFNYGLSNPYTNLNLTAAYNSGRQYKPKSGKFTGWGLLKDFETYNNAKILD 420
QY 420 NNTATFRLPRETELOTTGLGFNYFHNEYKGNRPPEELGLFFDGPDOODNGLYSLGRFKDGK 479
Db 421 NNTATFRLPRETELOTTGLGFNYFHNEYKGNRPPEELGLFFDGPDOODNGLYSLGRFKDGK 480
QY 480 GLLPQKSTIVPAGSQYFNTFYDAALKKDIYRLNYSNTVYRGFGGYTGYGSDDEFK 539
Db 481 GLLPQKSTIVPAGSQYFNTFYDAALKKDIYRLNYSNTVYRGFGGYTGYGSDDEFK 540
QY 540 RAGENSPTYKKHCNRSQGIYEPVLKYYGKKRANNSHSVISADFGDYFMPFASYSRTHRM 599
Db 541 RAGENSPTYKKHCNRSQGIYEPVLKYYGKKRANNSHSVISADFGDYFMPFASYSRTHRM 600
QY 600 PNQEMYFSQIGDSGVHVALPERANTWQFGNTYKGLLKODDPLGLKLVYRSRIDNY 659
Db 601 PNQEMYFSQIGDSGVHVALPERANTWQFGNTYKGLLKODDPLGLKLVYRSRIDNY 660
QY 660 IHNVYCKWMDLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFELBNLDYGRFFTNLSYA 719
Db 661 IHNVYCKWMDLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFELBNLDYGRFFTNLSYA 720
QY 720 YKSTQPTNFSASPNNAKEDOLKQGYGLSRVSALPRDYGRLEVTGRLGNKLTLLG 779
Db 721 YKSTQPTNFSASPNNAKEDOLKQGYGLSRVSALPRDYGRLEVTGRLGNKLTLLG 780
QY 780 AMRYFGKSIRATAEERYIDGTNGNTSNFROLGKRSIKQETTLARQPLIFDFYAAAYEPK 839
Db 781 AMRYFGKSIRATAEERYIDGTNGNTSNFROLGKRSIKQETTLARQPLIFDFYAAAYEPK 840
QY 840 NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLCKNGKYGT 899
Db 841 NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLCKNGKYGT 900
QY 900 KSVLTNFARGRTFLMTMSYKF 921
Db 901 KSVLTNFARGRTFLMTMSYKF 922

RESULT 3
G64110
hypoetical protein H1217 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: G64110
R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: G64110
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

QY 142 DSNFIAGLDVVKSGSAGINSLAGSANLRTLGVDVVOGNNYTGLLKGLGTSTNKG 201
Db 259 DPDLVSSVTKGPGSTRSAGI---GGSEVMRTIGVKDILVDGKDLGVRFTGDMVNN---G 312
QY 202 NAMAAGARKWLSSGASV---GVLYGHSRR--SVAQNVYRVGGGQHIGNFGAEYLERR 254
Db 313 VAQHSASAKTENLSPVHDDRGSLFGSOAKSGSAFAIR---NEHL-DLVAAYAQRN 367
QY 255 KORYFVOEGALKNSDPSGKWERDLQROQWKYKPKYNNQOE---LQKYEIHD----- 304
Db 368 QGNVYF-----SGKKQD-----RYRVNRYGREESSVAKYVYNAAGEEVLNSSSE 410
QY 305 -----KSHRENLAQYDIT-----PIDSSLKQOAGNLFKLEYGV-FNKY 345
Db 411 TSYLLKATWR--IADEHLDLQYRRYDGTGEIMPDSIFRGTAGTAYQYPLSEVKIDTY 468
QY 346 TAQFRLNTKIGSKIKINRYQFNYGLSNPYNLNTAAYNSGROKYPKSKFTGWGLL 405
Db 469 TARYRYPEN-----NPLVDLSTGLWTEAKSDMLTSLVAPRSQAYRSDRWNT----- 516
QY 406 KDPETYNNAKI-LDLNNTATFRL-----PRETELOTTLGFN----- 440
Db 517 ----RQDNRRIGGLANVARFETDFGDKLDLGGSFQVEDIQPKSVVTTLHDINARTL 572
QY 441 -YFHEYKGN-----RPEELGLFDDG-----PDQDNGLYSLGRPKGKGLLPKSTI 488
Db 573 RDAATREYGLNGKLEKPKVERLTWGGGRYSHFNKSDNGISAPRREDRDMRFI---TV 628
QY 489 VQPA---GSOYF---NTFYDAA--LKKDIYRLNTSTNTVGYRFGGEYTYGYGSDDEFK 539
Db 629 SRPGYGSMMWFQDNGQYTDATDPRLNGIVNTNINNPEGIFP-----DEFG 677
QY 540 RAGENSPTYKKHCNRCGIIYEPV-LKKYOKKRANHNHVSISADP---GDYFMPFASYSR 595
Db 678 PA-----NVTVHPSRVNTVYNTYKSGSRGGGFSFAPGFINFELAPOTFV-YASYTE 730
QY 596 THRPNIQEMFYSGIDSGVHTA--LKPERRANTWQGFNTYKGLLKQDDTLGKLVGYR 653
Db 731 GLRLPSLFE---TSQGTQVPEKDKLPERSRSEWEGASALRSLDAGDSAAIKLAYFN 787
QY 654 SRIDNTYHNHYGKWDLNGDIPSWSVSTGLAYTIQHRNEKDKVHKHGFLELNYDYGRFF 713
Db 788 NTIKNYITRYD-----PGOMGLMTFSNT-----DSYRTSGLEQLQSHYDAGRVF 831
QY 714 TNLAYAQKSTQPTNPSDASESPNNASKEDQLKQYGLSRVSAL-----PROYGLR 764
Db 832 ADLSATYYLKTETCDAAFAARLAGANRYORTENTPCTPGSMGSYTNTQNPRLATNL 891
QY 765 EVGTRWLGKLTLLGGAMRYFGKSIRATAERYIDGTNGGNTSNFROLGKRSIKQETFLAR 824
Db 892 TAGLRFEDQALTLLGRMTY-TSGPTATADKPQVG-----ATTPQIEYRSVQ----- 937
QY 825 QPLIFDYFAAYEPKKNLIFRAEYKKNLFRDRYIDPLDAGNDAATQRYYSFDPDKDDEVT 884
Db 938 ---LFDLFYKLFHEHTELNASLQNLTDYLYDPL-----AQSEMP----- 975
QY 885 CNADKTLNCKYCGTSKSVLTNFARGTFLITMSYKF 921
Db 976 -----APGRTLVRGMQAKF 989

RESULT 7
AH0477
probable TonB dependent receptor protein YPO3923 [imported] - Yersinia pestis (strain CC
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AH0477
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, T.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AH0477
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-830 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC93388.1; PID:g15981834; GSPDB:GN00175
C;Genetics:
A;Gene: YPO3923

Query Match 8.6%; Score 419.5; DB 2; Length 830;
Best Local Similarity 21.8%; Pred. No. 2.4e-18;
Matches 211; Conservative 138; Mismatches 347; Indels 271; Gaps 41;

QY 40 LEDVHVAKRVPKDKK-VFTDARAVS--TRODIFKSSSENLDN-----IVRSIPGAF 88
Db 50 LDKLNVGKGNADSDWIYDEPRSVSEITR-----EQLDNRPARHAADILEQTPGVYS 102
QY 89 QQDKSSGIVSLNIRGDSGFRVNTWVDGITQTFYSTDAGRAGSSQFGASVDSNFIAG 148
Db 103 SVSQQDPGLSINIRGIQDYGRVNMNIDMRQNMKSG--HGQRNGSMY----IDPEILSN 156
QY 149 LDVVKSGSFGSAGINSLAGSANLRTLGVDVVOGNNYTGLLKGLGTSTNKG--GNAMAA 206
Db 157 VTIERGIFNGIGGAGAGGATATFNTINASDFLAPEKELGGHIRAMTGDNGTRFTGSGALA 216
QY 207 IGARKWLESASGVLYGHRSRVSVAQNVYRVGGGQHIG-----NFGAEYLERRKQRY 258
Db 217 LG-----NPNGDILLAVSERNL-KDYWPGNKGVLAGRLYSPTRNVGDD-LKNTKTLF 267
QY 259 FVQEGALKFNSDSCKWERDLQROQWKYKPKYNNQOELOKYIEGHDKSWRENLAPOYDIT 318
Db 268 -----TGKMRSQLAKGVWNEFAGORLEFSYLQTOIAS-----PNASMLSEVL 310
QY 319 PIDPSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR-----KIIRNYOFTNGLS 373
Db 311 ALSPSGKE-----ITKIGWNTSFTNVNERNIALDYRLN 344
QY 374 LNPYTNLNLTA--YNSGROKYPKSGKFTGWLKDF-----ETYNNAKILDLNNTATRLP 428
Db 345 PEHISWLDATAKIYYVDNDETNDAN-----SLFKEYFTQTRLKTRGLQONTNTPPS 399
QY 429 RETELOTTLGFNFHNEYKGNRPEELGLFDDGPDQNGLYSYLGRFKGDKGLLPQK--- 485
Db 400 DAHQIRLKYGLEWF-----SDKSEG---YSTRKLIERTTPGKRAI 437
QY 486 -STIVQAGSOYFNTFYFDDAALKDIYRLNYSNTVYRFGGEYTYGYSDDEKRAFGE 544
Db 438 TSTFAQ-LNVEYDDWLREGLGRYDQFRLKGTWLHTRSFLOPY----- 480
QY 545 NSPTYKKHCNR-----SCGIYEPVLKKGKRRANHH-----SVSISADFG-DYFMP 589
Db 481 ---TFENPCDRRIHQSEKPGSRCSRPPATMRWDVDRCEQQLSPTLAMGYKPGLEWLEF 537
QY 590 FASYSRTRHPNIOEMFYFQIGDSGVHTALKP-----ERANTWQGFNTYKGLLKQDD 643
Db 538 FGSYGKSWRPAMTEVLAT--GTAHGYSWLPNPFVAERARTWEAGFNIOOSNLFIEDD 595
QY 644 TLGLKLVGYSRIDNYTHNYVYKWM-DLNGDIPSWSVSTGLAYTIQHRNEKDKVHKHGF 702
Db 596 HFAAKVAYFDTRIANYINLELGRAPKFGGD-----SFTDVAYV-----NNLLKTRFRGLE 646
QY 703 LELNYDYGREFTNLS-----YAYOKSTQPTNPSDASESPNNASKEDQLKQY 749
Db 647 YQLSYDAGTFYTNITNTRMTGVNVCSPYAWLGGLOSKYKGVKQVBOIYAVENEVANNY 706
QY 750 -----GLSRVSAALPRDYGRLEVGTRWLGKLTLLGGAMRY-FGKSTRATAEERYID 798
Db 707 VTCMANVLEFGSS--AYLPDGRGSLTIGSFIDRLDFGVIRYKNGYQDSAQDE---- 760
QY 799 GTNGGNTSNFROLGKRSIKQETLARQ--LIFDYAAYEPKKNLIFRAEYKKNLFRDRYI 856
Db 761 ---NGN-----PLTAYVADWPKYIVDYLYASYKVTNNLILRSSTENTINRAYL 805

QY 857 DPLDAGNDAATQRYSSDFDPKDKDEDTVCNADKTLNGLNGKYGTGSKSVLTNFA--RGRTFL 914
Db 806 -----VNYGDR-----LSFAPSRGRTIQ 823
QY 915 ITMSYKF 921
Db 824 GGFYKF 830
RESULT 8
C98310
hasR protein (Y08983) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: C98310
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: C98310
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-852 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK90005.1; PID:g15159974; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2875
A:Map position: linear chromosome
Query Match 7.3%; Score 359; DB 2; Length 852;
Best Local Similarity 21.1%; Pred. No. 1.6e-14;
Matches 200; Conservative 117; Mismatches 336; Indels 296; Gaps 37;
QY 62 AVSTRQDIFKSS--ENLDNIVRSIPGAFQTQDKSS--GIVSLNIRGDSGFGVRNVTWVGITQ 119
Db 111 SVVSREAIQAGVRNTRDFNVRVSGYAGGNGSFPTSPNVRGLOESGRVVSIDGARQ 170
QY 120 TF---YSTSTDAGRAGSGSQFGAS-----VDSNFIAGLDVVKGSFSGSAGINSLAGSANL 171
Db 171 NAORGYST-----GGASIYSANNGQAYVDAAFIRAVEVEKMTSATSNAGSLGKVEF 223
QY 172 RTLGVDVVOGNTYIGLLKGLTGTN--STKGNAAMAIAGKWLKESGASGVLYGHSRRS 229
Db 224 RTVSAADLIPGANKGAENVNVRSGNGYDFQGSVLAAV-----REPDGPLSFVAGYS-RT 277
QY 230 VAQNYRVGGGGQHI-----GNFG---AEYLERRKQRYFVQEGA 264
Db 278 IMDEYKIGTKGKALSTALTMKDLLGRDGSWTFPKSGDFGVDVQTSLSWMHQNDVFOGAS 337
QY 265 LKFNDSGK-----WERDLQROQWKYPKYN--NOELQKYTEGHDKSWRENLAPO 314
Db 338 TVVDREGVRNDSIVAKLMDWDPESLIDFKSSLWLDNMWTHLR-----AART 384
QY 315 YDI-TPID-----PSSLKQOASAGNLFKLEYDGVFNKYTAQ--FRDLNTRKIGSKRIINRY 366
Db 385 YAVETNLDMLGRSFGSLENTSR--FDTRAGALSILNYGAEAPRDIATSVATSAIQNP 441
QY 367 QF--NYGLSLNPYTNLNTAAYNSGRQPKPKSGKFTG-----WGLLKDFETYNNAKILDL 419
Db 442 SFASSY-TSFSAPAGRRDVASLFLNGELEPADWTLLSGGVRYDMSRLKGSATYYSFK-----496
QY 420 NNTATRLPRETELQTLGFTNYFNHNEYGNRPPEELGLFFDGDQDNGLYSLGRKFGDK 479
Db 497 -----ESIVTTSVPCDLVRNH-----513
QY 480 GLLPQKSTIVQAGSYFNFTFYDAALKKDIYRLNYSNTNTVYRFGGYTGYGSDDEFK 539
Db 514 -----TALEYFNQVFLPANLPVWASRYNVFLASIW-----543
QY 540 RAFGENSPYKKHCNCSGIEYPLVKKYKKRANNHSVLSADFG-----DY 586
Db 544 -----PRTSANCMPGTGI-----TTKTPVTEYPSHEVDIDRTYSAWLPSATIEFKPVDW 592

QY 587 FMFPASYSRTHRPNIQEMYF--SQIGDS-----GVHTALKPERANTWQGFNTYKKGLL 639
Db 593 FRPYVYSQSLRPTTILETFAGARPQDSAGYEYAPNOSLRAEKATYIEIGANKSFDGVL 652
QY 640 KQDDTLGLKLVGYSRIDNVI-----HNVYCKWDLNGIDIPSWSSSTGLATYIOHRNF 692
Db 653 LDDDTLRIRMAAFRREVKDYIALGYLVTQVDFRTY-----TSFVNLDDGTTY-----699
QY 693 KDKVHKHGFELNLYDYGREFTNLSYAYOKSTOPTNFSDASESNNASKEDQLKQGYGLS 752
Db 700 -----MRGELEBNYDARSFWIGSATVLKTEWPEKTQVFSNSTTTTSGE-----744
QY 753 RVSALPRDYGRLEVGTRWGLNKLTLGAMRYFGKISRATAERYIDGTNGGNTSNFRQLG 812
Db 745 -IVAWPGD-----VAPKM---KLTLDDGMRFFDEKESLGRNLNHNVTPTQSRTLDTGCLN- 794
QY 813 KRSIKOTELAROPILDFDYAAVEPKNLFRAEVKNLFRRIYIDPLDAGNDAATQRYYS 872
Db 795 -REITDPYTTV-----DLYGSYAFNDKATLREAVNNLTDRKYIPAASA-----YT 838
QY 873 SFPDPKDKDEDTVCNADKTLNGLNGKYGTGSKSVLTNFAARGRTFLITMSYKF 921
Db 839 -----APGRTFIATMNVPK 852
RESULT 9
AH2972
heme receptor hasR [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AH2972
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
; Karp, P.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH2972
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-923 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAL44198.1; PID:g17741777; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: hasR
A:Map position: linear chromosome
Query Match 7.3%; Score 359; DB 2; Length 923;
Best Local Similarity 21.1%; Pred. No. 1.8e-14;
Matches 200; Conservative 117; Mismatches 336; Indels 296; Gaps 37;
QY 62 AVSTRQDIFKSS--ENLDNIVRSIPGAFQTQDKSS--GIVSLNIRGDSGFGVRNVTWVGITQ 119
Db 182 SVVSREAIQAGVRNTRDFNVRVSGYAGGNGSFPTSPNVRGLOESGRVVSIDGARQ 241
QY 120 TF---YSTSTDAGRAGSGSQFGAS-----VDSNFIAGLDVVKGSFSGSAGINSLAGSANL 171
Db 242 NAORGYST-----GGASIYSANNGQAYVDAAFIRAVEVEKMTSATSNAGSLGKVEF 294
QY 172 RTLGVDVVOGNTYIGLLKGLTGTN--STKGNAAMAIAGKWLKESGASGVLYGHSRRS 229
Db 295 RTVSAADLIPGANKGAENVNVRSGNGYDFQGSVLAAV-----REPDGPLSFVAGYS-RT 348
QY 230 VAQNYRVGGGGQHI-----GNFG---AEYLERRKQRYFVQEGA 264
Db 349 IMDEYKIGTKGKALSTALTMKDLLGRDGSWTFPKSGDFGVDVQTSLSWMHQNDVFOGAS 408
QY 265 LKFNDSGK-----WERDLQROQWKYPKYN--NOELQKYTEGHDKSWRENLAPO 314

Db	409	TVVDREGVRNDSIVAKLDWDPESELIDFKSSSLWLNNDNMTHEL	-----AART 455
Qy	315	YDI-TPID-----PSSLKQOAGNLFKLEYDGVFNKYTAQ-FRDLNTRKIGSRKKIIRNY	366
Db	456	YAVETNLDMLGRSFGSLENTSR--FDTRAGALSINYGAEAFRIATSVATSIAQNP	512
Qy	367	QF--NYGLSLNPTNLNLTAAVNSGRQKYPKGSKFTG-----WGLLKDFETYNNAKILDL	419
Db	513	SFASSY-TSFSAGRRDVASLFLNGELEPADWITLGGVRYDMSRLKGSATYYSFK----	567
Qy	420	NNATFRLPRETELQTLGFNFHNEYGNRPPEELGLFFDQDQDNGLYSYLGRFRKSKDK	479
Db	568	-----ESIVTTSVPCDLVRNHY	-----584
Qy	480	GILLPQKSTIVQAGSYNFTFYDAAALKDKIYRLNYSNTVGYRFGGEYTYGYGSDDEFK	539
Db	585	-----TALYFNQVLEPANLPVWASRYNVFLASIW-----	614
Qy	540	RAFGENSPYKKHCNRSCGIYEPVLKKYKKRANHVSISADFG-----DY 586	
Db	615	-----PRTSANCMPGTGI-----TTKTPVTEYPSHEVIDIRTYSAWLPSATIEFKPVDW	663
Qy	587	FMPFASYSRTHRMNIOEMF--SOIGDS-----GVHTALKPERANTWQGFNTYKKGLL	639
Db	664	FRPYYSQSLSRPTTILETFAGARPGDSAGYEYAPQSLRAEKATYIEGANMSFDGVL	723
Qy	640	KQDPTGLKLVYRSIDNYI-----HNYGKWDNLNGDIPSWVSSTGLAYTQHNF	692
Db	724	LDDTLRIKMAAFRRVKDYIALGLVLTQDQVDRTY-----TSFVNLDGTTY-----	770
Qy	693	KDKVHKHGFELNLDYDGRFTNLSYAYQSTQPTNFSASESPNASKEDQLKQYGLS	752
Db	771	-----MRGLEGNFYDARFVWGGSTVLUKTEWPKTQVFSNSTTTTSGE-----	815
Qy	753	RVSALPRDYGRLEVGTRWLGKNTLGGAMRYFGKSTRATAERYIDGTNGGNTSNFRQLG	812
Db	816	-IVAWPGD-----VAPKM--KLITDGMRFDEKESLGRNLNHNVTPTQSRITLDTGNL-	865
Qy	813	KRSIKQTETLARQPLIFDYAAVEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATORYYS	872
Db	866	-REITDPYTV-----DLGYSYAFNDKATLRFVAVNNLTDRKIYPAASA-----YT	909
Qy	873	SFDPKDKDEDVTCNADKTLGCKYGGTSKSVLTNFAFGRTFLITMSYKF 921	
Db	910	-----APGRFTIATMNVKF 923	
RESULT 10			
D82437			
TonB receptor-related protein VCA0625 [imported] - Vibrio cholerae (strain N16961 serogroup O1; strain N16961; biotype El Tor)			
C:Species: Vibrio cholerae			
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001			
C:Accession: D82437			
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;			
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.			
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.			
Nature 406, 477-483, 2000			
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.			
A:Reference number: A82035; MUID:20406833			
A:Accession: D82437			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-784 <HEI>			
A:Cross-references: GB:AE004392; GB:AE003853; NID:g9658031; PIDN:AAF96526.1; GSPDB:GN001			
A:Experimental source: serogroup O1; strain N16961; biotype El Tor			
C:Genetics:			
A:Gene: VCA0625			
A:Map position: 2			
Query Match 6.3%; Score 309.5; DB 2; Length 784;			
Best Local Similarity 21.2%; Pred. No. 1.8e-11;			
Matches 195; Conservative 122; Mismatches 343; Indels 261; Gaps 43;			

Qy	6	RLKPCIFYLMGVTLHYHYAEADAGRAGSEAIQVLEDVHVAKRVPKDKKVFDTDAVST 65	
Db	36	QLKPL-FTLLPVVLSVVQAE-----NTEQAVDETIVH-----GOSILTDQR---T 79	
Qy	66	RODIFKSSNLDNIVRSIPGAFQODKSSGIVS-----LNRGDSGGGRVN 111	
Db	80	RSLDK-----VRGIANA-----DIFSGITSVQNNHNEAGALDGRVGOGEGRVP 127	
Qy	112	TWVDGTTQFYSTSDAGRAGSGSFQASVDSNFIAGLDVVKGS-----FSGSAGINS 164	
Db	128	IFDGLSQ---SHTSRGYQGVSDR--TYIDTLLSLTVNKGATIESSPYASGAVG--- 179	
Qy	165	LAGSAMLRLQVDDVVQGNNTYGLLKG-----LTGTNSTKGN-AMAAIGARKWL 213	
Db	180	--GVVNATTLGIKDIKDQAFGVWLKARANHNRTPDVSGDYSEOGQVALDERGHSFA 237	
Qy	214	ESGA-----SVGVLYGHSRRSVAQNYR-----VGGGGOHIGNFGAEYLE 252	
Db	238	KHGLMLGLGYOAESNTVLAYSKRSGKNHFAKKGYEYQEPVVGQGVVMTSE--- 294	
Qy	253	RRKORYFVOEGALKFNSDSGKWERDLQROQWKYKPKYKNYNNQELQYIEGHDKSWRENLA 312	
Db	295	-----SDSWLFKLASDTG---TAHNADFNVRHHAQKAGEVLMAYWYKSSDWEGNPY 343	
Qy	313	PQYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRKIGSKRIINRYQFNGL 372	
Db	344	PD-----GKDRMPQWGLGT-----AKVNTYSA-----NYYIQPDH-- 373	
Qy	373	SLNPTYNLNLTAAYNSGRQKYPKSGKFTG-WGLLKDFETYNNAKILDLNNTATFLPRET 431	
Db	374	---PWNLNANFWYTTADL-----AQNLGWLALGTNAEQYFAY---HNDRSGLSTNET 422	
Qy	432	EL-QTTLGFNYHNEYGNRFPPEELGLFDDQDQNGLYSYLGRFGDKGLLPKQSTIVO 490	
Db	423	LLTQWPVRLNYGLAQNERLSPEE---DQGT-----RF-----TKTITS 458	
Qy	491	PAGSQYNTFYFDAALKKDIYRLNYSNTVYRFGGEGTYGYGSDDEFFKRAFGENSEPTYK 550	
Db	459	RHGKRTAQNLFANADIDYSPRLVOLGLNLHNAK-----STDYQTKOOL-----DYK 504	
Qy	551	KHCNRCGIYEPVLKGYKKRANNSVSIADFGDYFMP---FASYSRTHRPNIQEMY 606	
Db	505	E-----KLDLSEFTYALTPTQFLKSSRYRPMPSLYETT 540	
Qy	607	FS-QIGDSGVHTALKPERANTWQGFNTYKKGLLKODDTLGLKLVGYRSRIDNYIHNVYG 665	
Db	541	LSNEVFSYNPNFIKPEQAWNNEVGQVFMASNSVLQDDRLNLSVSYFRNSIKDFISG--G 598	
Qy	666	KWDDLNGDIPSWVSSTGLAYTIQHRNFKDKVHKHGFELNLYDYGRRFTNLSTAYQKSTQ 725	
Db	599	RLAKTPG-MSEW--QANFTET---NY-DKLQISGWELGAHYQYAWLYTHFAATLYSETK 650	
Qy	726	PTNFSASESPNNAKEDOLKQCYGLSRYSALPRDYGRLEVGTRWLGKNTLGLGAMRYE- 784	
Db	651	ICSVQQAQYAESDTC--NSLGFAMGLTPTRIPPKQNLNLYNVGKFFNDTLDSGVKVSISHS 708	
Qy	785	GKSIRATAEERYIDGTNGGNTSNFRQLGKRSIKQTETLARQPLI-----FDYAAEYP 837	
Db	709	GKS---NPSDWLAGT-----AANPILEIPSDYIDLYSOYEL 742	
Qy	838	KKNLIFRAEVKNLFDRIYIDP 858	
Db	743	NANTOLFFAINNVTDRYQVRP 763	


```
Db 464 KSDRMIEESRNLQAVFKKAFDTAKIRHNLINLGYDRFKSQLSHSDYILQNAVQAYDL 523
Qy 507 -----KKDIYRLNYSNTVG-----YREGGE-YT-----GYGSDDEFFK 539
Db 524 ITPKKPPPPNGSKDNPYRVRSIGKTTVNTSPICRFGNNTYTDCTPRNIGGNGYAAVQDNV 583
Qy 540 RAFGENSPTYKKHCNRSGIYEPVLKKYKKRANHVS-----ISADFGDIYEMPFAZY 593
Db 584 R-----LGRWADVAGIRYDIRSTHSEDKSVSTGTHRNLSWNAAGVVLKPFWTM 631
Qy 594 SRTH-----RMPNIQEMIFSOIGDSGVHTALKPERANTWQFG-----FNT 633
Db 632 DLTYRSTGFLPSPFAEMYGWRAGESLKTLDLPEKSFNREAGIVFKDGFNLEASYFNN 691
Qy 634 YKGLLKQDDTLGLKLVGYRSRDN-----YIH-----NVYK--WBDLNG 672
Db 692 AYRDLI-----AFGYETRTQNGQTSASGDPGYRNAQNARIAGINILGKIDWHGVWG 742
Qy 673 DIPSWSVSTGLAY-----TIQRNFKDKVHKHGFELNLDY--GRFTNL 716
Db 743 GLPDGLYST-LAYNRIRKVKDADIRADRTFVTSYLFDAVQPSRYVLGLGYDHPDGIWGINT 801
Qy 717 SYAYOKSTQPTNFSASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNKLT 776
Db 802 MFTYSK-----AKSVDEL-----LGSQAL 820
Qy 777 LGGAMRYFGKSIRATABERYIDGTNGGNTSNFROLGRKRSIKQTETLARQPLIFDFYAAE 836
Db 821 LNG-----NANAKKAASRR-----TRPWYVTVTDVSGYN 848
Qy 837 PKKNLIPRAEVKNLFDRIYI 856
Db 849 IKHLLTRAGVYNLLNRYV 868
```

Search completed: July 24, 2002, 08:54:48
Job time: 539 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen-Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 09:00:10 ; Search time 20.76 Seconds
(without alignments)
1717.760 Million cell updates/sec

Title: US-09-762-926-6

Perfect score: 4894

Sequence: 1 MRSSRLKPCIFYLMGVTLX.....SVLTNFARGRTFLITMSYKF 921

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2505	51.2	913	1 YC17_HAEIN	P45114 haemophilus
2	273.5	5.6	744	1 HXCI_HAEIN	P44523 haemophilus
3	273.5	5.6	915	1 TBPI_NEIGO	Q01996 neisseria g
4	269	5.5	908	1 TB12_NEIMB	Q06987 neisseria m
5	256.5	5.2	911	1 TB11_NEIMB	Q09056 neisseria m
6	249	5.1	1013	1 HGBA_HAEIN	Q9Kiv2 haemophilus
7	243.5	5.0	999	1 HGPE_HAEIN	O87296 haemophilus
8	242.5	5.0	943	1 LBPA_NEIMB	O67379 neisseria m
9	239.5	4.9	993	1 HGBC_HAEIN	Q9Kiv0 haemophilus
10	239.5	4.9	1063	1 HGPI_HAEIN	P44795 haemophilus
11	239	4.9	944	1 LBPA_NEIMA	Q9Jtk4 neisseria m
12	238.5	4.9	912	1 TBPI_HAEIN	P44970 haemophilus
13	233	4.8	723	1 Y262_HAEIN	P44600 haemophilus
14	228.5	4.7	1077	1 HGPA_HAEIN	Q9za21 haemophilus
15	227	4.6	999	1 HGPD_HAEIN	P44809 haemophilus
16	225	4.6	725	1 HXC2_HAEIN	P45357 haemophilus
17	209	4.3	810	1 HPUB_NEIMC	P96949 neisseria m
18	208	4.3	810	1 HPUB_NEIMA	Q9Jwa2 neisseria m
19	206	4.2	1094	1 HPF3_HAEIN	P44836 haemophilus
20	204.5	4.2	1067	1 HGBB_HAEIN	Q9Kiv1 haemophilus
21	201.5	4.1	1066	1 HGPC_HAEIN	Q9x442 haemophilus
22	196	4.0	1046	1 HHUA_HAEIN	Q48153 haemophilus
23	182.5	3.7	999	1 HGBA_HAEIN	Q57408 haemophilus
24	181	3.7	972	1 HGBB_HAEIU	Q47957 haemophilus
25	180	3.7	774	1 FECA_ECOLI	P13036 escherichia
26	179.5	3.7	676	1 HMUR_VERPE	Q56989 versinia pe
27	178.5	3.6	614	1 BTUB_ECOLI	P06129 escherichia
28	178.5	3.6	697	1 HEMR_YEREN	P31499 versinia en
29	176.5	3.6	1567	1 ICEN_XANCT	P18127 xanthomonas
30	175	3.6	700	1 YNCD_ECOLI	P76115 escherichia
31	173	3.5	972	1 HGBA_HAEIU	Q47952 haemophilus
32	172.5	3.5	710	1 FOXA_YEREN	Q01674 versinia en
33	165.5	3.4	614	1 BTUB_SALTY	P37409 salmonella

ALIGNMENTS

```

RESULT 1
YC17_HAEIN
ID YC17_HAEIN STANDARD; PRT; 913 AA.
AC P45114;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable tonB-dependent receptor H11217 precursor.
GN H11217.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
[1]
SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RL Influenzae Rd.,"
RL Science 269:496-512(1995).
[2]
IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.,"
RL Electrophoresis 21:411-429(2000).
CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U32801; AAC22870.1; -
CC TIGR; H11217; -
DR InterPro: IPR000531; TonB_boxC.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
DR Outer membrane; Receptor; Signal; TonB box; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 913 PROBABLE TONB-DEPENDENT RECEPTOR H11217.

```

```

34 163 3.3 1104 1 COLA_CLOPE
35 162 3.3 1258 1 ICEN_ERWHE
36 157 3.2 1322 1 ICEA_PANAN
37 156 3.2 815 1 FPVA_PSEAE
38 155.5 3.2 729 1 FHUE_ECOLI
39 153 3.1 809 1 PUPB_PSEPU
40 151.5 3.1 687 1 VIUA_VIBCH
41 146.5 3.0 1148 1 ICEK_PSEEX
42 145.5 3.0 1113 1 N116_YEAST
43 144.5 3.0 989 1 SERA_PLAAG
44 142 2.9 730 1 DCOR_LACS3
45 142 2.9 1034 1 ICEN_PANAN

```

```

P43153 clostridium
P16239 erwinia her
P20469 pantoea ana
P48632 pseudomonas
P16869 escherichia
P38047 pseudomonas
Q00964 vibrio chol
Q30611 pseudomonas
Q02630 saccharomyc
P13823 plasmodium
P43099 lactobacill
Q47879 pantoea ana

```



```

Qy 133 GSSGASVDSNFTAGLDVYKSGSGSAGINSLAGSANLRLTGLVDDVVOGNNYTGILLKG 192
Db 117 -YQGSIFPELRLRRVTVDRKNTSPQVNGGFGAGTKVFKTDARDFLQENQKIGGFLK- 174
Qy 193 LTGNS-----TKGMAAAGARKWLESAGSVGLYGHRSRVAQNYRVGGGQHIGNFG 247
Db 175 -YGNNSNNOKTYSTALVQNEQNIDL-----LLFGSVRN--AGDYKRPDMSKIL----- 222
Qy 248 AEYLERRKQRYFVOEGALKFNDSGKWERDLQROOWKYKPKYNNQBELQYIGHDKSW 307
Db 223 ---FSKNKQK---TGLIKLN-----WQISP-----EHLTLSSVYGIHKGW 257
Qy 308 RENLAPQYDITPIDPSSLKQOSAGNLKFLKLEVDGVFNKYTAQFRDLNTKIGSRKIINRYQ 367
Db 258 -EPAAKRDIIP-----KPSLDIMRYGTD--TAKRKLVIYD-----QKDNITYLK 301
Qy 368 FNYGLSLNPYNLNLTAAYNSGRQK--YPKGSKFTGMGLL--KDFERYNNAKILDLNNTA 423
Db 302 YNYPENNPWNLSTQFSYKTTQNDMRPKKASSGLVSLGNQSWITYSDL--TFDINNNTS 360
Qy 424 TFRLPRETELQTTLGFNYPHNEYKKNRFPPELGLFFDQPDQDNGLYSLVGRFKDGKGLLP 483
Db 361 TFNI-KTTVHELLFGLQWLNKT--RNT-----LMYDKSVRKADYNY-GYFQ----- 403
Qy 484 QKSTIVOPAGSQYFNTFYFDAALK-KDIYRLNYSTNTVGYRFGGEYTYGYGSDDEKRAF 542
Db 404 ---PYMPSGRQYTAQFYLDQIQKWNII---FST---GVRY-----DHINNIG 443
Qy 543 GENSTPYKHKHNSCGIYEPVLKYGKRRANNHS--VSISADFGDYFMPFASYSRTHRMP 600
Db 444 QKNLAL--KYNDISAG-----HDYSQNYNGWSYVGLNVDVNHLSLFTNFSKTWRAP 495
Qy 601 NIQEMYSQIGDSGVHTA---LKERANTWQFGNTYKKGKLLKQDDTLGLKLGVRSRID 657
Db 496 VIDEOYETQFOKSSVPATSLNEKEMINQTRVGGIITLHNFQENDAPQFETTYFYNRGK 555
Qy 658 NYIH-----NYVGKWLNDLNGDI--PSWSSSTGLAVTIOHRNEKDKVHKHGFLELNYDGR 711
Db 556 NEIFKTRGVNCVGNADNNKVCPIIEN-----YRNLPGYVIO-GAELEYQSTY 606
Qy 712 FFTNLAYAQSTOPTNFSADASESPNNASKEDQLKQGYLSRVSAALPRDYGRLEVGTRWL 771
Db 607 LFGEITYSVYKGRDT-----SPRN-----PWGKTSTWIAEIPPRKATATGALGNVP 652
Qy 772 GNKLTGAMRYFGKSIPTAEERYIDGTNGNTSNFQPKRSIKTKTETLARQPLPDF 831
Db 653 KYITVWGWAEEVRRQDRSPL-----SGDPKASSWSLPASRGYS-----CHNL 695
Qy 832 YAAYPEK--KNLIFRAEVKNLFDTRY 855
Db 696 FLSWSPAKIKGMVNIITVDNLFNRAY 721

RESULT 3
TBPL_NEIGO STANDARD; PRT; 915 AA.
AC Q01996;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA19.
RX MEDLINE=92394880; PubMed=1325963;
RA Cornelissen C.N., Biswas G.D., Tsai J., Paruchuri D.K.,
RA Thompson S.A., Sparling P.F.;
RT "Gonococcal transferrin-binding protein 1 is required for transferrin
utilization and is homologous to TonB-dependent outer membrane

```

```

RT receptors.";
RL J. Bacteriol. 174:5788-5797(1992).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M96731; AAA25503.1;
DR PIR; A43335; A43335.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 915 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB BOX.
FT SITE 898 915 TONB C-TERMINAL BOX.
FT SITE 915 915 MW: 97CF74B1010422F CRC64;
SQ SEQUENCE 915 AA; 102213 MW; 697CF74B1010422F CRC64;

Query Match 5.6%; Score 273.5; DB 1; Length 915;
Best Local Similarity 21.3%; Pred. No. 1.3e-09;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

Qy 5 FRLPDPICVLMGVTLHYHYAEDAGRAGEAQIQVLEVDHVHAK--RVPKDKKVTTDARA 62
Db 7 FRLNLICLSMTAL---PAYAENV-QAG-QAQEKOLDITQVAKKQKRRDNEVTGLKGL 61
Qy 63 VSTRQDIPKSFSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGRYNTWVDGIT 118
Db 62 VKTADTLKEQVLDLRDLTRYDPGIAVVEQGRGASSG---YSIRG-MDKNRVSLVVDGLA 117
Qy 119 Q-TFYSTSTDAAG--RAGSSQFAGSVDSNFTAGLDVYKSGSGSAGINSLAGSANLRTIG 175
Db 118 QIQSYTAQAALGGTGTAGSSGAINIEIENVKAVEISKGSNVSVEOGSAGALAGSAVFQTK 177
Qy 176 VDDVVOGNNYTGILLKGLTGTNSTKGNMAAIGARKWLESAGSVGLYGHRSRVAQNYRV 235
Db 178 ADDVIGEGRWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212
Qy 236 VGGG-----GOHIGNFGAEYLERRKORYFVQEGALKFN-----SDSGKWERDLQROOW 283
Db 213 IGGAEALLIRTGTHAGEITRAHEAAGR-----GVQSFNRLAPVDGSKYAYFVEEE- 263
Qy 284 KYPKYKYNNOEL-----QKYLEGHDK---SWRENLAQYDITPIDPSSLKQOS----AG 331
Db 264 ---CKNGGHEKCRANPKKDVVEDKQTVSTRDYTGNNRFLA--DPLSYESRSWLFRRPG 317
Qy 332 NLF--KLEYDGVFNKYTAQ---FRDLNTKIGSRKII-----NNRYQFN--YG 371
Db 318 FRFENKRHYIGGILLERTQOTEDTRDMTPVAFITKAVFDANQKQAGSLRGNGKYAGNHRYG 377
Qy 372 LSLNPYTNLNTAAYNSGRQYKPKGSKFTGMGLLCKLDFETYNNAKILDLNNTATFRLPRET 431
Db 378 -----GLFTSGENNAVPAGAEY-GTGVFYFD-----ET 402
Qy 432 ELQTTLGFNYPHNEYKKNRFPPEELGLFFD--GPDDQDNL----- 468
Db 403 HTKSRYLEYVYVTRADKDTWADYARLSYDRQIGLDNHFQOQTHCSADGSDKYCRPSADKP 462
Qy 469 YSYLGRFKDGKGLLPQKSTIVQAGSQYNTFYFDAALKKDIYRLNYSTNTVGY-RFGG- 526
Db 463 FSY---YKSDRVIVGESHKLLQAAPKKSFDTAKI-----RHNLNVN-LGYDRFGSN 509

```

QY 527 -----EYTGYYGSDDEKRAFGNSP----- 547
 Db 510 LRHODY--YIGSAN--RAYSLKTPQNGKKTSPNGREKNPYWVSIGRGNVTVTRQICLF 564
 QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YKGRANNHVS-----I 579
 Db 565 GNNTYDCTPRSINGKSYAAVRDNVRLGRWADVGAGLRYDYRSTHSDGSGVSTGTHRTL 624
 QY 580 SADFGDYFMPFASYSRTH-----RMPNIOEMYSQJGDSGVHTALPERANTWQFG--- 630
 Db 625 SWNAGIVLKPADWLDLRYSTGRFLSPFAEMYGWRSGDKIKAVKIDPEKSFNKEAGIVF 684
 QY 631 -----FNTYKKGKLLKQDDTLGLKLVGYRSRDN-----YIH----- 661
 Db 685 KGDFGNLEASWNNAYRDLIVR-----GYEAQKDGKEQVKGPNAYLNAQSARITG 735
 QY 662 -NYGK--W---WDLNGDIP--SWVSTGLAYT--IQHNFVKDKVHKHGFEELELNVDYGRFF 713
 Db 736 INILGKIDWGVND---KLPEGWYST--FAYNRVRVDIKKRDRTDIQSHL--FD----- 784
 QY 714 TNLASYAQKSTOPTNFSDASESPNASKEDQKOGYGLSRVSALPRDYGRLEVGTRWLG 773
 Db 785 -----AIQPSRYV-----VGSYDQPEGKGVNGMLT-----YSKAKEITELLS 824
 QY 774 KLTGGAMRYFGKSIRATAEERYIDTNGGNTSNFROLGKRSIKQETLARQPLIDFYA 833
 Db 825 RALLNG-----NSRNTKATARRTRPWVIVDVS 852
 QY 834 AYEPPKNLIPRAEYKNLFDRIYI 856
 Db 853 YTVKHKFTLRAGVYNLLNHRIV 875

RESULT 4
 TB12_NEIMB
 ID TB12_NEIMB STANDARD; PRT; 908 AA.
 AC Q06987;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transferrin-binding protein 1 precursor.
 GN Tbp1
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RC SEQUENCE FROM N.A., AND SEQUENCE OF 25-42.
 RX STRAIN-CCUG 37603 / B16B6 / SEROGROUP B / SEROTYPE 2A;
 MEDLINE-93345825; PubMed-8344530;
 RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
 RA Jacobs E., Schryvers A.B.;
 RT "Cloning and characterization of Neisseria meningitidis genes
 encoding the transferrin-binding proteins Tbp1 and Tbp2.";
 RL Gene 130:73-80(1993).
 RP [2]
 RN SEQUENCE OF 25-45.
 RC STRAIN-CCUG 37603 / B16B6 / SEROGROUP B / SEROTYPE 2A;
 RX MEDLINE-93307625; PubMed-8319886;
 RA Griffiths E., Stevenson P., Byfield P., Ala'Aldeen D.A.,
 RA Borriello S.P., Holland J., Parsons T., Williams P.;
 RT "Antigenic relationships of transferrin-binding proteins from
 Neisseria meningitidis, N. gonorrhoeae and Haemophilus influenzae:
 RT cross-reactivity of antibodies to NH2-terminal peptides.";
 RL FEMS Microbiol. Lett. 109:85-91(1993).
 CC -1- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
 TRANSFERRIN UTILIZATION.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- INDUCTION: BY IRON STARVATION.
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z15129; CAA78831.1; -
 DR PIR; JN0819; JN0819.
 DR PIR; S33154; S33154.
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_boxC; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Receptor; Signal; TonB box.
 FT SIGNAL 1 24
 FT CHAIN 25 908 TRANSFERRIN-BINDING PROTEIN 1.
 FT SITE 38 45 TONB BOX.
 FT SITE 891 908 TONB C-TERMINAL BOX.
 SQ SEQUENCE 908 AA; 101583 MW; FE2FF4974CAC3C31 CRC64;
 Query Match 5.5%; Score 269; DB 1; Length 908;
 Best Local Similarity 21.2%; Pred. No. 2.4e-09;
 Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;
 QY 5 FRLKPICYFLMGVTLHYHYSAEADAGRAGSEAQIQVLEDDVHVAK--RVPKDKVFTDARA 62
 Db 7 ERLNLCLSMTALPV---YAENV--QAEQAQEKQLOTIQVAKKQKTRRDNVETGLGKL 61
 QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDQK--SSGIVSLNIRGDSGFRGVRNTWVGIT 118
 Db 62 VKSSDTLSKEQVLNIRDLTRYDPGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGVS 117
 QY 119 Q-TFYSTSTDAG--RAGSSQFGASVDNFTAGLDVVYKGSFGSAGINSLAGSANLRTG 175
 Db 118 QIOSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKGSNSSEYNGALAGSVAFTQT 177
 QY 176 VDDVVGQNTYGLLLK-GLTGTNSTKGNMAAIGARKWLESAGSVGLYGHRSRVSQNY 234
 Db 178 AADIIGKQWGIQSKTAYSGDKHALTQSLALAGR---SGGAELLIIYTKRRGREHAH 233
 QY 235 RVGGGGQHGIFGAEXLERKQ-----RYFVQEGAL-----KFNSD-SCKWER-- 276
 Db 234 KDAGKG--VOSFNRLVDEDKEGGSQYRYFIVEECHNGYAAACKNKLKEDASVYDKERT 291
 QY 277 -----DLQOQWKYP-----YKNYNNQELQKYLEGHPKSWRENLAPO 314
 Db 292 VSTQDYTGSNRLANPLEYGSQWLFPRGWHLNDRHYVGAVLERTQQTFT--RDMTVPA 349
 QY 315 YDTPID--PSSLK-----QQSAGNLFKLEY-DGVFNKYTAQPRDLNTK 355
 Db 350 Y-FTSEDYVPGSLKGLGKYSQDNKAERLFVQGGEGSTLQIGIGYGVF-----YDERHTK 402
 QY 356 IGRKLIINRYQFNYSGLSNPYTNLTAAYNSGRQKYPKSGKFTGWGLLKDFETYNNAK 415
 Db 403 -----NR-----YGVVEY-VYHNADKDTWADYARLSDYDROG----- 431
 QY 416 ILDLNNTATFRLPRETELQTTLGFNYEHNEYKGNRFPPEELGLFPDQDNGLYSYLGRF 475
 Db 432 -IDLN-----RLQQT---HCSHSDGSKNCRP-----DG---NKPYSF--Y 463
 QY 476 KDKGLLPQKSTIVQAGSOFYNT-----FYFDAAL----- 506
 Db 464 KSDRMIEESRNLFQAVFKKAFDTAKIRHNLSINLGLYDRFKSOLSHSDYLLQNAVOAYDL 523
 QY 507 -----KKDIYRLVNSTVVG---YRFGGE-YT-----GYGSDDEFK 539
 Db 524 ITPKKPPFPNGSKDNPYRVSIGKTTVNTSPICRFGNNTYTDCTPRNIGGNGYAAVDNV 583
 QY 540 RAFPENSPYKKHCNRSKCIPEYVLKYGKGRANNHVS-----ISADFGDYFMPFASY 593
 Db 584 R-----LGRWADVGAGIRYDYRSTHSEKDSVSTGTHRLNSWAGVVLKPTWM 631
 QY 594 SRTH-----RMPNIQEMYSQJGDSGVHTALPERANTWQFG-----FNT 633


```
Qy 391 QYKPKGSKFTG-----WGLLKDFETYNNAKILDLNNTATFRLPRETELQTLTGFIYFNEY 446
Dy 347 WLAKLGYRFGGRHVGVGFED--TKQYDIDRM-----TEKQ-----YGGTDE 387
Qy 447 GKNRPEELGLFFGPDODNGLY--SVLGRPKGDKLLPKQKSTIVQAGSOYFNTFFEDA 504
Dy 388 AE-KFRQKSGV-YGDDDFRDLGLYFVPIEEWKDKNL-----VRGILKYSRTKFTDE 438
Qy 505 ALKKDIYRLNYSNTVVG--YRFGGEYTGYYGSDDEF-----KRAFGENSPYKKHKNRSCG 558
Dy 439 -----HRRRRMGLLYRYENEAYSDNWADKAVLSFDKQGVATDNTLTK-----LNCA 485
Qy 559 IYEPVLKYYKGRANNHVSISADFGYFMPFASYSRTHRPNTQMYFSO-----609
Dy 486 VY-PAVDKSCRASADK---PYSYSSDRF-----HYREGH-----NYLNASFEKSLKNKWTKH 534
Qy 610 -----IGDSGVHTALKPER-----ANTWQFGFNYYKGLLKQDDTLGLKLKLVY 652
Dy 535 HLTGFGVDASKAISRPQLSHNAARISESTGFENNQDKYLLKPEVVEGSGCYIETL 594
Qy 653 RSR--IDNYIH--NVY-----GKWD--LMG--DIPSVSSTGLA-----684
Dy 595 RSRKCVPRKINGSNIHSLNDRFISIGYFDFSLGRYDRKNFTTSEELVRSRYVDRSWN 654
Qy 685 -----YTIQHR---NFKDKVHKHGFELNYYD-----709
Dy 655 SGILFKPNRHFVSYRASSGFRPTSFQELGIDYHDPKQWRPALKSEKAAAREIGLQ 714
Qy 710 --GRF--FTNLVAYOKTPTNFSD--ASESPNNASKEDOL-----KQGYGLSRVSAIP 758
Dy 715 WKGFDFLEISFNRNYDMTAVADHKTKLPNQAGLQTEIDIRYNNQNNMSLOGVNLG 774
Qy 759 R-----DYGRLEVG--TRWLGKNTLGGAMRYFGKSIR-----ATAERYIDG-----799
Dy 775 KIDNWGVYGLPEGLYTTLAYNRKPKSVSNRPGLSRLSYALDAVQPSRYVLGFGYDOPE 834
Qy 800 -----TNGGNTSNFROL--GKRSIKQETELARQPLIFDFYAYEPKKNLIFRAEV 847
Dy 835 GKWGANTMLTVSKGNPDDELAYLAGDKRYSTKRASSWSWADVSAYLNLKRTLRAAI 894
Qy 848 KNLFDRIYIDPLGDADAATQRYYSFSDPKDKDEDVCNADKTLGKNGYKYGTSKSVLTNF 907
Dy 895 YNIGNYRV-----TWESLRQTAESTAN-----RHGDS-----NY 925
Qy 908 AR-----GRFTLITMSYKF 921
Dy 926 GRYAAPGRNFSALEMKF 943

RESULT 9
HGBC_HAEIN STANDARD; PRT; 993 AA.
AC QKIVO;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin binding protein C precursor.
GN HGBC.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTHI N182;
RX MEDLINE=20316037; PubMed=10858226;
RA Cope L.D., Hxal Z., Hansen E.J.;
RT "Detection of phase variation in expression of proteins involved in
RT hemoglobin and hemoglobin-haptoglobin binding by nontypeable
RT Haemophilus influenzae."
RL Infect. Immun. 68:4092-4101(2000).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND
```

```
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAGING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF221060; AAF80178.1; -
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC
CC Outer membrane; Transport; TonB box; Multigene family; Signal;
CC Receptor; Repeat.
CC SIGNAL 1 24 POTENTIAL.
CC CHAIN 25 993 HEMOGLOBIN BINDING PROTEIN C.
CC DOMAIN 26 37 3 X 4 AA TANDEM REPEATS OF Q-P-T-N.
CC REPEAT 26 29 1.
CC REPEAT 30 33 2.
CC REPEAT 34 37 3.
CC SITE 47 54 TONB BOX.
CC SITE 976 993 TONB C-TERMINAL BOX.
CC SEQUENCE 993 AA; 113616 MW; A551BF3B2C641612 CRC64;
CC
Query Match 4.9%; Score 239.5; DB 1; Length 993;
Best Local Similarity 18.4%; Pred. No. 1.9e-07;
Matches 214; Conservative 153; Mismatches 380; Indels 417; Gaps 50;
Qy 3 SSERLKPICFYLM-GVTLYHYSYAEDAGRAGSAQIQ-----VLEDVHVK-----AK 48
Dy 2 ANFRNLVAYSVMGLGTAGVAYAAQPTNQPTNQNSNVSEQLQNVSGSTENSSTK 61
Qy 49 RVPKDKVFTDARAVSTRQDIFKXSENLDNIVSRIPGAFTQODKSSGIVSLNIRGDSGF- 107
Dy 62 TPKIAETVKTAKTLERQ-----ANNIKDIVKYETGV-----TVVEAGRFQSGFA 108
Qy 108 -----GRVNTWVDGITQTFYSTSTDAGRAG-----GSSQFGASVDSNFIAGLDVVK 153
Dy 109 IRGVDENRVAINIDGLROA--ETLSSQGFKELFEGYGNFNNTNRNGAET--LKEVNITK 164
Qy 154 GSFSGSAGINSLAGSANLRTLGVDDVVOGNTYGLLLAGLTCTNSTKGNMAAICARKWL 213
Dy 165 GANSIKSGSGSLGGSVIYKTDARDYLKNDYVYKGYATENNQSFNTLTLAGRYKKF 224
Qy 214 ESGASVGLYGHRSRVSVAQNYRVGGGGQGHIGNFGAEYLERRKQRYFVQEGALKFNSSDGK 273
Dy 225 DA-----LVVYTRN-----GHELENY--DY---KNADSLTQKKRERADPVK 262
Qy 274 WERDLQROQWYKPYKNYN---NOELQKY-TEGHDKSWRENLAPOYDITPDPSSLKQOS 329
Dy 263 IEQDSTLLKLSFNPTENHFRFTLAADLYEHRSGQDLSY-----TLKYQR 306
Qy 330 AGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKLIINRYFNY-GLSLNPTNLTAAVNS 388
Dy 307 SGN-----ETPEVESRHT-----NDKTKRNRISFSYENFSQTPFWD-TLKLTYSD 350
Qy 389 GRQK-----YPKGSKFTGWGLLK-----D 407
Dy 351 QRIKTRARTDEYCDAGVTHCGGTENPTGLKLTNGKITRNGTDLQFKEKGTAAADGTKTYD 410
```



```
QY 250 YLERKORYFVOEGALKFNDSOGKWERD-----LQOQWQYKPKYKNYNNOEL 296
D 248 RLDAFRQTYDIQOKNKAYEFLABGERPKPVAKLAGNGYLLKQLNRWVEERKKNQPL 307
QY 297 QKYIEG---HDKSWRENLAPOYDIT---PIDPSSLKOOSAGNLFKLEY-----DGVF 342
D 308 NAEEMVREAAQAHENLSAQ-SYTGGRILPDPMDYRSGSLWAKLGYRFGGRHYVGCVF 366
QY 343 --NKYTAQFRDLNFK---IGSRKIILNRN---YQFNGLSNPY-----TNLNL 383
D 367 EDTQRVDIRDMTEKQYGYGDEAKFKSNKSGYDGNDFRGLYFVNPNEBKWGTNLVK 426
QY 384 AAYNSGRQKY-PKSGKFTGWLKDFE-----TYNNKILDLNNT----- 422
D 427 IGLKYSRTKFTDEHRRRRMGLLYRYENEAYSDNWADKAVLSFDKQGVATDNTKLNCA 486
QY 423 -----ATFERLP---RETELQTTLGFNFHN 444
D 487 VYPAVDKSCRASADKPKSYDSSDRFHYREQHNVLNASFESLKNKWKHHLTLGFGVDAS 546
QY 445 E-----YGNRPFPEELGFDGPDQDNGLY-----SYLGRFKGDKGLLPQ 484
D 547 KAISRPEQLSHNAARISESTG--FDDNNQDKYLLGKPEVVEGSGVGIETLSRK-CVPR 603
QY 485 K---STIVQAGSOYFNTFTDAAKKDIYRLNISTVTVGYRFGGEYTYGYGSDDEPKRA 541
D 604 KINGNIHISLINDSFISGKYGFDLSLGRYDRQNFTEELVRSG----- 647
QY 542 FGENSEPTYKHCNRSOG---IYEPVLKYGKKRANNH-SVSIADFGDYFMPFASYSRTH 597
D 648 -----RYVDRSWNSGIVFKP-----NRHFSVSRASSG-----F 676
QY 598 RMPNIOEY-----FSQIGDSGVHTALKPERANT-----WQ---FGF-----NTYKKGL 638
D 677 RTPSQFQELFGIDYHDYKQWOPALKSEKAANREIGLQWKGDFGLEISSFRNRYTDMI 736
QY 639 LKQDDTLGL-KLVGYRSRID--NYIH-----NYYGK--WDLNGDIPSWSSVSTGLA 684
D 737 AVADHKTLQLPDSTGRLTREIDIRDYNAQNMSLOGVNILGKIDWNGVYKLPGLYTT-LA 795
QY 685 YT-----IQHNEK-DKVHKGHFELELNYDY--GRFTNLSYAYQKSTQPTNF 729
D 796 YNRKPKSVSNRPDLRLSRVALDAVQPSRYVLGFGYDQPEGKWCANIMLYISKGNPDE- 854
QY 730 SDASESPNNASKEDQLKQGGKLSRVSALPRDYGRLEVGTRW-----LGNKLTLLG 779
D 855 -----LAYLAGDQKRYSTKRASSWSWSTADVSAYLNLAKRLTLRA 893
QY 780 AMRYFG-----KSIRATAEERYIDGTNGNTSNFRLQKRSIKQETTLARQLIFDFY 832
D 894 AIYNIGYRVVTWESLRQTAES---TANRHGGDSNYGR-----Y 929
QY 833 AAYEPKKNLIFRAEVK 848
D 930 AA--PGRNFSIALEMK 943

RESULT 12
ID TBPI_HAEIN STANDARD; PRT; 912 AA.
AC P44970;
DT 01-NOV-1995 (Rel. 32, Created)
DE 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable transferrin-binding protein 1 precursor.
GN TBPA OR HI0994.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
```

```
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32780; AAC22656.1; -
CC TIGR: HI0994; -
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box; Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 912 PROBABLE TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 50 57 TONB BOX.
FT SITE 895 912 TONB C-TERMINAL BOX.
SQ SEQUENCE 912 AA; 103233 MW; 294C08991A652CC2 CRC64;

Query Match 4.9%; Score 238.5; DB 1; Length 912;
Best Local Similarity 20.3%; Pred. No 2e-07;
Matches 215; Conservative 134; Mismatches 413; Indels 299; Gaps 50;

QY 5 FLRKPICFYLMG--VTLYHYSYAEDAGRAGEAQIQVLED-----VHVKAARV--PKDK 55
D 7 FLRSIIISCLLISCVRAETOSIKDKREAISSVEVDTOSTEDSELETISVTAEKVRDRKDNE 66
QY 56 VFTDARAVSTRQDIFKSSSE-NLDNIVRSIPGAFTQODKSSGIYSLNIRGDSGFRVNTMV 114
D 67 VTGLGKIITSEISIREQVLNIRDLTRYPDGLSVVQGRGASSGYSIRG-MDRNRVALLY 125
QY 115 DGITQT-FYSTSTDAGRAGSSQFGA--SVDSNFIAGLDVYVKGFSGASGINSLAGSANL 171
D 126 DGLPQTQSYVQSPVLVARSYSGTGAINETIENYKRAVEISKGSSEYNGALAGSVTF 185
QY 172 RTLGVDDVVOGNNYTLGLLKGLTGTNSTKG--NAMAAGARKWLESGASVGLVYHGRSRS 229
D 186 QSKSAADILEGDKSWGIOTKN-AYSSKNKGFTSLAVAG-----KQGGFGLAIYQTRNS 239
QY 230 VAQNYRVGG--GGQHGNGFCAEYLERKQRYFVQE----- 262
D 240 IETQVHKDALKGVSQSYNRLIAK--PENQSAIFYMEDECPKGYDECIPSAKPPAILSTKKE 297
QY 263 -----GALKFNSDGGKWERDLQROQWKYKPKYKNYNNOELQKYTEGHDKSWRENLAPO 314
D 298 TVSVSDYTGANRIKPNPKMYE---SQSWFLRGYHFSEQH---YIGGFIEFTQQ----K 346
QY 315 YDITPID-PSSLKQSQSAGNLF-----KLEY-----DGVFNKYTA--OFRDLNFK--- 355
D 347 FDIRDMTFPAYLSPTKEDGLANRPFPYKQDYGAYGHIEDGRGVKYASGLYFDEHRRKQRY 406
QY 356 -----IGSRKKIINRNYQFNGLSNPY---TNLNLTAAYNSGRQYKPGSKF 399
```

```

Db 407 GIEYIENKKNAGIIDKAVLSANQQ-----NIILDSYMRHTHCSL-----YPNPSKN 453
QY 400 TGMGLLKDFETYNNAKILDLNNTATFRLPRETELQT-----TLGFNYF-----HNE 445
Db 454 CRPTLDKPYSYSDRNVYKEKHNLQNLNLEKKIQIOWNLTHQIVENLGFDFETSAHQKD 513
QY 446 YGRNPFEEGLGDFD--GPDQDGL--YSYLGKPKGDKGLLPKOKSTIVQAGSQYFNTFY 501
Db 514 YLRRVIATADSIDTKGTRRGLREYPL-----YKPKPYF----- 552
QY 502 FDAALKKDIYRLNVTNTVYRGEGEYTGYSDDPEKRA-----FGENSPY 549
Db 553 -----AGEDHCNTQGSNNYRDCVKRLIKGKNYFFARNNMAL 590
QY 550 KKHNRSCGIYEPVLKYYKKRANHSVSI-----SADEGDFYMPF-----ASV--SRTH 597
Db 591 GKVDYDLGLGIRYDV-----SRTKANESTISYGVKPKNFESWNTGIVIKPPEWLDLSYRLSTGF 646
QY 598 RMPNIQIYFSQIG-----DSGVHTALKPERANTWQFGNTYKGLLKQDDPLGLKLGYRS 654
Db 647 RNPFSFAEYGMWYGGKNDYVYGVKFKPETSQNOEF-----GLALKGDFGNEISHFSN 699
QY 655 RIDNYIHNVYKGMWDLNDIPSWSSYGLAYTIOHRNFKDKVHKHKGPELELNDYDYGREET 714
Db 700 AYRNLI--AFAEELSKNG-----TGKNGYGHNAQNAKL--VGVNITQAQLDNFGLWK 747
QY 715 NLSYAYOKSTQPTNFSASESPNNASKEDQKQYGLSLRSYAL-----PRDY-----GRL 764
Db 748 RIPYGMW-----ATPAYNRVKVKQO--KINAGLASVSSYLFDAIQSPRYIIGLYD 796
QY 765 EVGTRWLGKLTGGAMRYFGKSRATAEERYIDGTNGNGNTSNFRQLGKRSIKOTETLAR 824
Db 797 HISNTW-----GVNATETQS--KAKSONELLGKRALGNNS-----RDVKSTRKLTR 840
QY 825 QPLIFDYAAEYKPKNLIFRAEYKNTFDREYIDPLDAGNDAATQYRYSFDPKDKDEDVT 884
Db 841 AWHILDVSGYMAKNMLRLGLYNLFENRY-----VT 873
QY 885 CNADKTLGNGYKGTGSKSVLTNFAR-----GRTEFLITMSYKF 921
Db 874 WEAVRQTAQGA--NQHQNVGNTRYAASGRNYTLTLEMKF 912
RESULT 13
Y262_HAEIN STANDARD; PRT; 723 AA.
AC P44600;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN HI0262.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Karlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Sequence 269:496-512(1995).

```

```

RN [2]
RX IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
CC -|- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
CC -|- SUBCELLULAR LOCATION: Outer membrane (potential).
CC -|- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32712; AAC21927.1; -.
CC TIGR; HI0262; -.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC.1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE; PS01186; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Receptor; Signal; TonB box; Complete proteome.
CC SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 723 PROBABLE TONB-DEPENDENT RECEPTOR HI0262.
CC FT SITE 706 723 TONB C-TERMINAL BOX.
CC FT SEQUENCE 723 AA; 80775 MW; 91EB3ABOFFEA2984 CRC64;
CC -----
Query Match 4.8%; Score 233; DB 1; Length 723;
Best Local Similarity 20.0%; Pred. No. 3.2e-07;
Matches 184; Conservative 133; Mismatches 311; Indels 292; Gaps 47;
QY 31 AGSEAQIQVLVDVHVAKRVP-----KDKVFTDARAVSTQDIFKSENLDNI 79
Db 17 ANALQSGVELDSINVIATRDPSRFAYTPERQSKDSLKSQATSV--DALEDIPNYD-- 71
QY 80 VRSIPGAFTQDDKSSGIVSLNIRGDSGFRVNTMVDGITQTFYSTSTSDAGRAGSSOFGA 139
Db 72 VRGGSRIAQKP-----NIRGLSD--NRVVQIDVGRQNF-----DLAHRG--SYF-- 113
QY 140 SVDSNFTAGLDVVKGFSFGSAGINSLAGSANRLTLGVDDVVGQNNYGLLLK--GLTGNS 198
Db 114 -LPMSLIQIEIVIKGPPSSLWGSGLGVVAMRTPNALDLLKNDKFGVKIROGYQTANN 172
QY 199 TKGNMAAIGARKWLESGASGVLYGHSRVSVAQNYRVGGQGHGFGAEYLERRKQRY 258
Db 173 LSEKDVSVFAANDKFD--VLISGFYNN-----ADNLTGKGNK-----LNNTAYKQ 216
QY 259 FVQEGALKFN---SDSGKWERDLQROQWKYKPKYKNYNQELQKYIEGHDKSWRENLAPOY 315
Db 217 F--GGLAKFGWINDANRVE--LSHRETRFKQTAPSNN-----EVENELTNEQITDOI 265
QY 316 -----DITPIDPSLKKQOSAGNLFKLEYDGVFNKYTAQFRLNLTGKS-----RKI 361
Db 266 KKFHGGKDDLLP--PTTQPPSPSERSEF-----YSKVKTRLGVSVSLYTDQOI 309
QY 362 INRNYQNYGLSL--NPYTNLNLTAAYNSGRQKYPKSGKFTGWLGLKDFETYNNAKILDLN 420
Db 310 PDOSTVFNYLTPDNPNYLNTHIALYNNKTIKEQOR--KVSG--VVD--QTKLATRGNLR 363
QY 421 NTATFRLPRETELQTLTGPNYFNEYGNKFRPELGFDFGPDQDNGLYSLVGRFKGDKG 480
Db 364 NSSEL-----SHISFVYGVDMRDKIKTER-----GTNG 392
QY 481 LLFQKSTIVQAGSQVFNTFYFDAALKKDIYRLNVTNTVYRGEGEYTGYSDDDEFKR 540
Db 393 -----SDAKFRADPY--NANSTTGVVLIHIP----- 418
QY 541 AFGEN---SPTYK--KHONRSCGIYEPVLKYYKKRANHSVS-----ISADFGDYFMPFA 591

```

```
Db 419 LFGEKLVSPSVRYDHYDTS-----SKTVKYKDNHLSPATKLTWIVTNWLDL---TA 467
QY 592 SYSTRHPNIOEMFYQIGDSGVHTA-----LKPERRNTWQFGENTY 634
Db 468 KYNEAFRAPSQMERFV-----SGAHFGANTGLDHNFRVANPLRPETAKNKEITANLH 522
QY 635 KGLLKQDDTLGLKLVGRSRIDNYIH-NVY-----GKWDNLNGDIPSWVSTGLA 684
Db 523 FDSLFGKQDFKFIENYFRNDVDFINLKIFENDAKTSASAGANPTINGAL-----LP 574
QY 685 YTHQRNFKDKVHKHGFELNIDYGR--PFTNLSYAYQKSTOPTNFSDESPPNASK 742
Db 575 KNSOYQNTN-ARLSGIELQAQYQTERLTFTNY-----GSTKG 612
QY 743 DQLKQGVGLSRVLSALPRDYGRLEVGTRW--LGNKLTILGGAMRYFGKSIRATAERYIDGT 800
Db 613 KDKDSGEALSNIAA-----SKIGVGNVALVKDFVGATVTHYAAQ-----654
QY 801 NGGNTSNFRLGKRSIKQETELARQPLIFDYAAYEPK-----KNLIFRAEVKNLFDORRYI 856
Db 655 -----RRVPRKDHSTVTPSYLTDLRATYAPLKGWKNLRLDPALENLFDORRYQ 702
QY 857 DPLD-----AGNDAATQRYYS 872
Db 703 PAFSLMEGTGRNAKISAVYS 722

RESULT 14
HGPA_HAEIN
ID HGPA_HAEIN STANDARD; PRT; 1077 AA.
AC G9ZA21; 09R649;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin and hemoglobin-haptoglobin binding protein A precursor
DE (Heme-repressible hemoglobin-binding protein) (Hgb).
GN HGPA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HI689 / Serotype B;
RX MEDLINE=9923576; PubMed=10220170;
RA Jin H., Ren Z., Whitby P.W., Morton D.J., Stull T.L.;
RT "Characterization of hgpa, a gene encoding a haemoglobin/haemoglobin-
RT haptoglobin-binding protein of Haemophilus influenzae.";
RL Microbiology 145:905-914(1999).
RN [2]
RP SEQUENCE OF 1-145 AND 988-1077 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=HI689 / Serotype B;
RX MEDLINE=96333350; PubMed=8757844;
RA Jin H., Ren Z., Pozsgay J.M., Elkins C., Whitby P.W., Morton D.J.,
RA Stull T.L.;
RT "Cloning of a DNA fragment encoding a heme-repressible hemoglobin-
RT binding outer membrane protein from Haemophilus influenzae.";
RL Infect. Immun. 64:3134-3141(1996).
RN [3]
RP ROLE OF CCAA NUCLEOTIDE REPEATS.
RX MEDLINE=99412292; PubMed=10482534;
RA Ren Z., Jin H., Whitby P.W., Morton D.J., Stull T.L.;
RT "Role of CCAA nucleotide repeats in regulation of hemoglobin and
RT hemoglobin-haptoglobin binding protein genes of Haemophilus
RT influenzae.";
RN [J. Bacteriol. 181:5865-5870(1999)].
RC CL -/- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE.
CC -/- SUBCELLULAR LOCATION: Outer membrane.
CC -/- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
```

```
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAINING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -/- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: U51922; AAD10835.1;
CC InterPro: IPR000531; TonB_boxC.
CC Pfam: PF00593; TonB_boxC; 1.
CC PROSITE: P500430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE: P501156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Transport; TonB box; Multigene family; Signal;
CC Receptor; Repeat.
CC SIGNAL 1 24
CC CHAIN 25 1077
CC
CC HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN
CC BINDING PROTEIN A.
CC
CC DOMAIN 26 69
CC 11 X 4 AA TANDEM REPEATS OF P-T-N-Q.
CC REPEAT 26 29 1.
CC REPEAT 30 33 2.
CC REPEAT 34 37 3.
CC REPEAT 38 41 4.
CC REPEAT 42 45 5.
CC REPEAT 46 49 6.
CC REPEAT 50 53 7.
CC REPEAT 54 57 8.
CC REPEAT 58 61 9.
CC REPEAT 62 65 10.
CC REPEAT 66 69 11.
CC SITE 78 85
CC TONB C-TERMINAL BOX.
CC SITE 1060 1077
CC TONB BOX.
CC SEQUENCE 1077 AA; 122813 MW; 693F673BB5AC59F1 CRC64;

Query Match 4.7%; Score 228.5; DB 1; Length 1077;
Best Local Similarity 19.5%; Pred. No. 1e-06;
Matches 211; Conservative 131; Mismatches 351; Indels 389; Gaps 48;

QY 40 LEDVHVYKARVPKDKVFTD-ARAVSTRQDIFK-SSENLDNIVRSIPGAFQODKSSGIV 97
Db 77 LEQINVSGSTENTDTRAPPKIAETVKTAKLEQAQDVKDLVRYETGI-----TVV 128
QY 98 SLNIRGDSGF-----GRVNTWVDGITOTFYSTSTDAG-----RAGSSQFGASVDSN 144
Db 129 EAGRFNSGFVRGVVEENRVAVOIDLHQA--ETISQGFKELFEGYGNFNTRNAAEIE 186
QY 145 FTAGLDVVRKGSFSGSAGINSAGLSANLRTLGVDVVGQNNYTGLLKLTGTNTSGNAM 204
Db 187 TLKQVTRIRKADSLKSGSALGGSVSLDTKARDYLLNKNYASYKRGYNTADQNQLNTL 246
QY 205 AIGARKWLESAGSVGLYGHRSRVAQNVYRGGGQGHIGNFG-----AAYLERK 255
Db 247 TLGGRYKYFD--AAI AVL--TSRK-----GHELENFGYKNYNDKIQGTREKAD 290
QY 256 QRYFVQEGAL-----KENSQSGKWERDLQROQWKY--KP--YKNYNNQELQKY 299
Db 291 PYRTQDSALLKIGFQPTNHRFSVVADLYKQTSKGHDFSYTLKPNQTYWYDEKELR-- 348
QY 300 IEHDKSWRENLA-----POYDITPDPSSLKQOS-----AGNLF 334
Db 349 -HTNDRVERKNIAFVYENFETFPWDTLKITYSHQITTSARTDDYCDGNDKCALAGNPL 407
QY 335 KLEY-----DGVFNKYTAQFDRLN-TKIGSRKLIINRYQFNGLSNPTNLT 384
Db 335 KLEY-----DGVFNKYTAQFDRLN-TKIGSRKLIINRYQFNGLSNPTNLT 384
```

```

Db 408 GMYNODNQLVGKDG-----KSAKYQDINKTQVIKERL-----PFTKPN--- 446
QY 385 AYNNGROKYPKSGKFTGWLK-----DFETYN----- 412
Db 447 ----GRWRFHK-----VDWDALKKKYGPVPIYASCLEEDNDPSEFCYEVKTKKENTFEI 498
QY 413 NAKILDNLNTA-----TFLPRET-----ELQTLGPNFYFN----- 444
Db 499 NGARYDLLSEADKNVLSDEQRPLTNVSYLFSQDGLNCKDKTILGFKKRRNLLKIFLFEVI 558
QY 445 ----EYGNRFPPEELGFLFDGPDQDNGLYSYLGRKFGDKGLL-----SQY----- 496
Db 559 EKRCQYGTQVKYKAN-----DQLSGPYLFMPNKKGYQANLWSQDLTSETKQINLQDL 610
QY 483 ----PQKSTIVQAG-----SQY----- 496
Db 611 TKHLELGKTHDLSYGLWSEMKSMNTNLAGDPTPLNVKWAQYPHNCATFLPDPSTWTPNA 670
QY 497 ----FNTYFEDAAL-----KDIYELNLYSTNVTGYRFG-----GE 527
Db 671 KPTLNPERTSTLCNNVNVFSFLIPVKTKTGALYFINDRVNNYVAFNLGYRYDRVKYEPE 730
QY 528 Y-----TGYGSDDEFFKRAFGNSP--TYKKHCNRSCGIYEPVLKKYGGKRA 572
Db 731 YIPGKTPKIPDDMTNLYIKTPEFASKADSDPDLSKKEANAANKETAQPK--KPSA 788
QY 573 NNHSVISADFGYFMPFASYSRTHMPNIQEWYFS-QIGDSGV--HTALKPRANTWOF 629
Db 789 SSYFSGTTLDPNLWLRLQAKYSGFRAPTSDEIYFTFKHPDFSQPNRDLQPETAKTEL 848
QY 630 GFNTYKGLLKQDDTLGLKLVGYSRIDNVIHNVYCKWDLNGD---IPSWYSSTGLAYT 686
Db 849 SLTVH-----NDMGYITTSFDTRYQNFIDLSYQGRDRDVHGSKLIPF----- 891
QY 687 IOHRNFKDKVHK-----HGFELNLYDYGR-----FFTNLSYAYQKSTQPTNFSDAE 734
Db 892 ----HYQNVNRNPAKVTGFEIASQISLGNITKLFNGFSLSYKYTYQKGRINGNIPMNAI 947
QY 735 SPNASKEDQKQYGLSRVSALPRDYGRLEVTRWLNKNTLGGAMRVFGKSIRATAEE 794
Db 948 QPRTAV-----YGVSVYH--PDDKYGLDL-----YISHASAKNAED 981
QY 795 RYIDGTNGGNTSNFROLGKRSIKQETLARQPLIEDFYAAEPKKNLIPRAEVKNLFDRR 854
Db 982 TY-----NMWYKEEGKTD-STIKWRSKSYTTIDLLGYKPIKNTLRAGVYNLTNRK 1032
QY 855 YI 856
Db 1033 YI 1034

RESULT 15
HGP2_HAEIN STANDARD; PRT; 999 AA.
ID HGP2_HAEIN
AC P44809;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 2
DE precursor.
GN HI0661.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty J.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

```

```

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.,
RA Peterson J., Hickey E., Dodson R., Gwinn M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
RN [4]
RP CONCEPTUAL TRANSLATION.
RA Coudert E.;
RL Unpublished observations (AUG-2001).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO
CC AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; TWO FRAMESHIFTS WERE
CC INTRODUCED IN THE REPEATS REGION AND IN THE LEADER PEPTIDE TO
CC MAXIMIZE THE SIMILARITY WITH OTHER ORTHOLOGS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32749; AAC22319.1; ALT_SEQ.
CC TIGR; HI0661;
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE; PS00156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Transport; TonB box; Multigene family; Signal;
CC Receptor; Repeat; Complete proteome.
CC SIGNAL 1 24 POTENTIAL.
CC CHAIN 25 999 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-
CC HAPTOGLOBIN BINDING PROTEIN 2.
CC DOMAIN 26 53 7 X 4 AA TANDEM REPEATS OF Q-P-T-N.
CC REPEAT 26 29 1.
CC REPEAT 30 33 2.
CC REPEAT 34 37 3.
CC REPEAT 38 41 4.
CC REPEAT 42 45 5.
CC REPEAT 46 49 6.
CC REPEAT 50 53 7.
CC SITE 63 70 TONB BOX.
CC SITE 982 999 TONB C-TERMINAL BOX.
CC SEQUENCE 999 AA; 114690 MW; 1A17AAB220092B7D CRC64;
SQ

```


Query Match 4.6%; Score 227; DB 1; Length 999;

Best Local Similarity 19.3%; Pred. No. 1.2e-06;

Matches 206; Conservative 146; Mismatches 398; Indels 320; Gaps 45;

QY 40 LEDVHYK-----AKRYPKDKVFDARAVSTRODIFKSSSENLNDNIVRSIPGAFQQDK 92
DB 62 LEQINVSSTENSDTKTPPKIAETVKTAKTLEREQ-----ANNKIDIVKETGV----- 110
QY 93 SSGIVSLNIRGDSGF-----GRVNTMVDGITQTFYSTSTADAGRAG-----GSSOF 137
DB 111 --TVVEAGREGQSGFAIRGVDENRVAINIDLQQA--ETLSSQGFKELEPGYGNFNNTNRN 166
QY 138 GASVDSNFIAGLDVWKGSGAGINSAGSAMLRLGVDVVOGNNTYGLLLKGLTGTN 197
DB 167 GAEIET--LKEVNTKAGDSIKKSGSLGSGSVIYKTKDARDYLINKDYVYVSKKYATEN 224
QY 198 STKGNAMAAIGARKWLESAGSVGLYCHSRVSVAQNYRVGGGGOHIGNFGAELERRKOR 257
DB 225 NQSFDTLTLAGRYKKFD-----VLVYVTSRN-----GHELENYGKKNYNDKIQ- 267
QY 258 YFVQEGALKFNSDSGKWERDLRQOMKYPKYNY-----NQOELQ---KYIE 301
DB 268 -----GKKREKADPYKIEQDSTLLKLSFNPTENHRTFAADLYEHRSGODLSYTLKYOR 322
QY 302 G-----HDKSWRENLAPOYD----- 316
DB 323 SGNETPEVDSRHTNDKTKRNNISFSYENFSOTPFWDTLKLTYSQRIKTRATDEYCDAG 382
QY 317 -----ITPIDPSLKOOSAGNLFKLE---YDGVFNKYTAQFRLNT 354
DB 383 VRHCEGTDPNTGLKVNGKTRRDGSDLOPEEKNNTAKSSDKTYD--FKK---FIDTDK 436
QY 355 KIGSRKIINRYOFNYGLSLNPTNLNTAAVNSGRQKYPKSGKFTGWG-----LLK 406
DB 437 RVIDDKLVLNN-----PSDTWDCSIFNCENNAKIKVFKGNNYGYDGKWKKEVDLEIK 489
QY 407 DFTYNNNAKILDLNNATFRLP-----RETELQT-----TLGFNYFHNE-- 445
DB 490 ELNGKFKAFIKNDRIKSIKILSPSPGYLERLWQERDLDTNTQOLNLDLTAKDFKIWHIEHN 549
QY 446 --YG-----KNRFPPEELGLFFDGPD-----QDNGLYSVLGR-----FKDGKGLLPQ 484
DB 550 LQYGGSYNTAMKRVNRAG--NDASDVQMWATPFLGEDSWTGKPHTCATTYEWANLCPR 607
QY 485 ---KSTIVPAGSQYFNTFYDAALKDIYRLNYSNTVNTYRFGG-EYTCYYG----- 533
DB 608 VDPEFSYLLPIKTKGKSVLYFDNFVITD-----YLSFDLGYRDNIHYQPKYKHGIPKL 662
QY 534 SDDEFKRAF-----GENSPYKHKHNRSCGIYEPVLKYKGRANNHVSISADFGDYFM 588
DB 663 PDDIVKGLFIPLPNNSNDPNKVENVQONI-DYIAKQNKYKAHSYFVSTIDPTSFLR 721
QY 589 PFASYSRTHRMPNIQEMYSF-QIGDSGV--HTALKPERANTWQFGFNTYKGLLKODDTL 645
DB 722 LQLYSKGFRTPTSDEMYTFKHFDPTILPNTDLKPEIAKTEIAFT-----LHNDWDG 775
QY 646 GLKLVGRYRIDNYIHNVYKMWDLNGDIPSWYSSTG---LAYTIOHRNFKDKVHKHGF 701
DB 776 FISTSLFKTYKNFIDLIFFKKQETFK-----VGGSGRGETLPFSLYQINIRDNASLKI 829
QY 702 ELELNYDYGR-----FFTNLSYAYOKSTQPTNFSASESPNNASKEDQLKOGYGLSRV 754
DB 830 EINSKVFLGKMAKMGDFNLKYTYQKGRMNGNI-----PMNATQPRTMVYGLGYDHP 883
QY 755 SALPRDYGRLVCTRWLGNKLTILGGAMRYFGKSIRATAEERYIDGTNGGNTSNFRQLGKR 814
DB 884 N-----HKFGDFDYTT-----HVASKNPEDTYN--NFYKE 911
QY 815 SIKQETL---ARQPLIFDFYAAEYKKNLIFRAEYKKNLFDRIYIDPLDAGNDAAATORY 871
DB 912 ENKKDSTIKWRSKSYILDLIGYVQPIKNLTIRAGVYNLTNRKYI-----TWSARSIRSF 967
QY 872 SSFDPKDKDEVTNADKTLNCGKYGGTSKSVLTNFARGRTFLITWSYKF 921

Db 968 GTSNVIDQ-----STGLGINRRFYAPGRNYKMSYQFEE 999
Search completed: July 24, 2002, 09:00:22
Job time: 413 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:59:42 ; Search time 68.49 Seconds
(without alignments)
2326.301 Million cell updates/sec

Title: US-09-762-926-6
Perfect score: 4894
Sequence: 1 MRSSRLKPICFYLMGVTLV.....SVLTNFARGRTFLTMSYKF 921

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4883	99.8	921	16 Q9JR50	Q9JR50 neisseria m
2	4763.5	97.3	922	16 Q9JTN4	Q9JTN4 neisseria m
3	1096.5	22.4	755	16 Q9PIV1	Q9PIV1 campylobact
4	901	18.4	753	2 Q93052	Q93052 shigella fl
5	900	18.4	733	2 Q93354	Q93354 escherichia
6	468.5	9.6	885	16 Q92NX4	Q92NX4 rhizobium m
7	435.5	8.9	883	2 Q9XSP4	Q9XSP4 pseudomonas
8	435.5	8.9	891	16 Q9HVJ7	Q9HVJ7 pseudomonas
9	428.5	8.8	989	16 Q9HUR6	Q9HUR6 pseudomonas
10	426	8.7	916	2 Q9RHT4	Q9RHT4 pseudomonas
11	412	8.4	899	2 P72473	P72473 serratia ma
12	357.5	7.3	848	16 Q9CKJ4	Q9CKJ4 pasteurella
13	346	7.1	747	16 Q98L70	Q98L70 rhizobium l
14	311	6.4	782	2 Q93SH4	Q93SH4 bradyrhizob
15	309.5	6.3	784	16 Q9KLM6	Q9KLM6 vibrio chol
16	279	5.7	739	2 O87381	O87381 haemophilus

17	272.5	5.6	907	2	Q9KJ54	Q9KJ54 neisseria g
18	271	5.5	912	2	Q9KK96	Q9KK96 neisseria g
19	269.5	5.5	693	2	Q56644	Q56644 vibrio chol
20	269.5	5.5	915	2	Q9KK97	Q9KK97 neisseria g
21	268.5	5.5	698	16	Q9KM13	Q9KM13 vibrio chol
22	266	5.4	912	2	Q9KI56	Q9KI56 neisseria g
23	265	5.4	791	16	Q9JYA8	Q9JYA8 neisseria m
24	265	5.4	910	16	Q53348	Q53348 neisseria m
25	264	5.4	791	2	P72084	P72084 neisseria m
26	262.5	5.4	688	2	Q93SS7	Q93SS7 plesiomonas
27	261.5	5.3	791	2	P72085	P72085 neisseria m
28	258	5.3	883	2	O53136	O53136 neisseria m
29	255.5	5.2	915	2	Q9JP30	Q9JP30 neisseria m
30	254.5	5.2	915	16	Q9K0U9	Q9K0U9 neisseria m
31	249	5.1	940	2	O51187	O51187 neisseria m
32	248.5	5.1	742	16	Q9CN63	Q9CN63 pasteurella
33	247	5.0	792	2	Q9RGS9	Q9RGS9 neisseria m
34	246	5.0	784	16	O9CMS1	O9CMS1 pasteurella
35	246	5.0	792	2	O51104	O51104 neisseria m
36	244.5	5.0	716	2	Q9AJS1	Q9AJS1 vibrio angu
37	242.5	5.0	743	16	Q92N43	Q92N43 rhizobium m
38	239.5	4.9	764	2	O68881	O68881 pseudomonas
39	239.5	4.9	764	16	Q9HV88	Q9HV88 pseudomonas
40	232.5	4.8	943	2	O50952	O50952 neisseria g
41	230.5	4.7	912	2	O48037	O48037 haemophilus
42	229	4.7	708	2	Q9XBV1	Q9XBV1 porphyromon
43	228	4.7	911	2	Q48044	Q48044 haemophilus
44	227.5	4.6	867	2	Q9KIB1	Q9KIB1 porphyromon
45	227.5	4.6	943	2	O87343	O87343 neisseria m

ALIGNMENTS

RESULT 1

Q9JR50
ID Q9JR50 PRELIMINARY; PRT; 921 AA.
AC Q9JR50
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE TONB-DEPENDENT OUTER MEMBRANE RECEPTOR (TONB-DEPENDENT RECEPTOR).
GN TDFH OR NMB1497.
OS Neisseria meningitidis, and
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487, 491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IR1074;
RA Turner P.C., Thomas C.E., Stojiljkovic I., Elkins C., Kizel G.,
RA Ala Aldeen D.A., Sparling F.P.;
RT "Identification and Characterization of Putative TonB-dependent Outer
RT Membrane Proteins among Pathogenic Neisseriae";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RL MC58";
RL Science 287:1809-1815(2000).
DR EMBL; AF227418; AAF73907.1; -;
DR EMBL; AE002499; AAP41853.1; -;

DR TIGR; NMB1497; -
DR InterPro; IPR000531; TonB_boxC.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 921 AA; 104222 MW; 67985B75EDF8819D CRC64;

Query Match 99.8%; Score 4883; DB 16; Length 921;
Best Local Similarity 99.8%; Pred. No. 4.6e-259;
Matches 919; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRSSFLRPICFYLMGVTLHYHYAEADAGRAGSEAOIQVLEHVHAKRVPKDKKVFDTA 60
Db 1 MRSSFLRPICFYLMGVTLHYHYAEADAGRAGSEAOIQVLEHVHAKRVPKDKKVFDTA 60

Qy 61 RAVSTRQDIFKSSSENLDNIVRSIPGFTQDQKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Db 61 RAVSTRQDIFKSSSENLDNIVRSIPGFTQDQKSSGIYSLNIRGDSGFRVNTWVDGITQT 120

Qy 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV 180
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV 180

Qy 181 QGNNTYGLLLKGLTGNTSTKGNAMAAIGARKWLESGASVGLYGHRSRVAQNYRVGGGG 240
Db 181 QGNNTYGLLLKGLTGNTSTKGNAMAAIGARKWLESGASVGLYGHRSRVAQNYRVGGGG 240

Qy 241 QHIGNFGAEYLERRKQRYFVOEGALKENSQKWERDLQROQWKYPKYKNYNNQELQYI 300
Db 241 QHIGNFGAEYLERRKQRYFVOEGALKENSQKWERDLQROQWKYPKYKNYNNQELQYI 300

Qy 301 EGHDKSWRENLAPOYDITPIDPSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNFKIGSRK 360
Db 301 EGHDKSWRENLAPOYDITPIDPSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNFKIGSRK 360

Qy 361 IINRNYQFNGLSLNPYTNLNTAAYNSGRQKYPKGSFKFTGWGLLKDFEYNNNAKILDLN 420
Db 361 IINRNYQFNGLSLNPYTNLNTAAYNSGRQKYPKGSFKFTGWGLLKDFEYNNNAKILDLN 420

Qy 421 NTATFRLPRETTELQTTLCGFNFHNEYGNRPPEELGLFFDQDQDNGLSYLGFRKGDGK 480
Db 421 NTATFRLPRETTELQTTLCGFNFHNEYGNRPPEELGLFFDQDQDNGLSYLGFRKGDGK 480

Qy 481 LLPQKSTIVQAGSQYFNTFYDFAALKKDIYRLNYSNTVTGYRFGGEYTYGYGSDDEFKR 540
Db 481 LLPQKSTIVQAGSQYFNTFYDFAALKKDIYRLNYSNTVTGYRFGGEYTYGYGSDDEFKR 540

Qy 541 AFGENSPYKHKCHNRSGIYEPVLKYYKKRANHHSVSISADFGDYFMPFASYSRTHRM 600
Db 541 AFGENSPYKHKCHNRSGIYEPVLKYYKKRANHHSVSISADFGDYFMPFASYSRTHRM 600

Qy 601 NIQEMYFSQIGDSGVHTALPERANTWQFGFNTYKGLLKQDDPFLGLKLVGYRSRIDNYI 660
Db 601 NIQEMYFSQIGDSGVHTALPERANTWQFGFNTYKGLLKQDDPFLGLKLVGYRSRIDNYI 660

Qy 661 HNVYTKWMDLNGDTPSWVSSTGLAYTTQHRNFKDKVHKHGFELNLDYGRFFTNLSYAY 720
Db 661 HNVYTKWMDLNGDTPSWVSSTGLAYTTQHRNFKDKVHKHGFELNLDYGRFFTNLSYAY 720

Qy 721 OKSQPTNFDSPASPNNAKEDQKOGYGLSRVSALPRDYGRLVGTWRGNKLTILGGA 780
Db 721 OKSQPTNFDSPASPNNAKEDQKOGYGLSRVSALPRDYGRLVGTWRGNKLTILGGA 780

Qy 781 MRYFGKSTRATAERYIDGTNGGNTSNFRQLGKRSIKQETLARQPLIFDFYAAEYEPKN 840
Db 781 MRYFGKSTRATAERYIDGTNGGNTSNFRQLGKRSIKQETLARQPLIFDFYAAEYEPKN 840

Qy 841 LIFRAEVKNLFDTRYIDPLDAGNDAATQRYYSFDPDKDEDVTCNADKTLGNGKYGGTS 900
Db 841 LIFRAEVKNLFDTRYIDPLDAGNDAATQRYYSFDPDKDEDVTCNADKTLGNGKYGGTS 900

Qy 901 KSVLTNFARGRTFLMTMSYKF 921
Db 901 KSVLTNFARGRTFLMTMSYKF 921

Db 901 KSVLTNFARGRTFLMTMSYKF 921

RESULT 2
Q9JTN4 PRELIMINARY; PRT; 922 AA.

ID Q9JTN4
AC Q9JTN4;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE PUTATIVE OUTER MEMBRANE SUBSTRATE BINDING PROTEIN.
GN NMA1700.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;
RA "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491";
RL Nature 404:502-506(2000).
DR EMBL; AL162756; CAB84928.1; -;
DR InterPro; IPR000531; TonB_boxC.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 922 AA; 104290 MW; B45E683985DE20B6 CRC64;

Query Match 97.3%; Score 4763.5; DB 16; Length 922;
Best Local Similarity 97.5%; Pred. No. 1.6e-252;
Matches 899; Conservative 9; Mismatches 13; Indels 1; Gaps 1;

Qy 1 MRSSFLRPICFYLMGVTLHYHYAEADAGRAGSEAOIQVLEHVHAKRVPKDKKVFDTA 60
Db 1 MRSSFLRPICFYLMGVTLHYHYAEADAGRAGSEAOIQVLEHVHAKRVPKDKKVFDTA 60

Qy 61 RAVSTRQDIFKSSSENLDNIVRSIPGFTQDQKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Db 61 RAVSTRQDIFKSSSENLDNIVRSIPGFTQDQKSSGIYSLNIRGDSGFRVNTWVDGITQT 120

Qy 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV 180
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV 180

Qy 181 QGNNTYGLLLKGLTGNTSTKGNAMAAIGARKWLESGASVGLYGHRSRVAQNYRVGGGG 240
Db 181 QGNNTYGLLLKGLTGNTSTKGNAMAAIGARKWLESGASVGLYGHRSRVAQNYRVGGGG 240

Qy 241 QHIGNFGAEYLERRKQRYFVOEGALKENSQKWERDLQROQWKYPKYKNYNNQELQYI 299
Db 241 QHIGNFGAEYLERRKQRYFVOEGALKENSQKWERDLQROQWKYPKYKNYNNQELQYI 300

Qy 300 IEGHDKSWRENLAPOYDITPIDPSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 359
Db 301 IEGHDKSWRENLAPOYDITPIDPSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 360

Qy 360 KIINRNYQFNGLSLNPYTNLNTAAYNSGRQKYPKGSFKFTGWGLLKDFEYNNNAKILDL 419
Db 361 KIINRNYQFNGLSLNPYTNLNTAAYNSGRQKYPKGSFKFTGWGLLKDFEYNNNAKILDL 420

Qy 420 NNTATFRLPRETTELQTTLCGFNFHNEYGNRPPEELGLFFDQDQDNGLSYLGFRKGDGK 479
Db 421 NNTATFRLPRETTELQTTLCGFNFHNEYGNRPPEELGLFFDQDQDNGLSYLGFRKGDGK 480

Qy 480 GLLPQKSTIVQAGSQYFNTFYDFAALKKDIYRLNYSNTVTGYRFGGEYTYGYGSDDEFK 539
Db 480 GLLPQKSTIVQAGSQYFNTFYDFAALKKDIYRLNYSNTVTGYRFGGEYTYGYGSDDEFK 539


```

Db 451 --KEGGFNSVQLSAQVTPWLQPPICGYSKSMRAPNIQEMFFNSGGASMPFLKPERAE 507
QY 626 TWQGFNTYKGLLKQDDTLGLKLVGRYSRIDNYIHN-----VYQKWD 669
Db 508 TWOAGFNIDRDLLEQDALFRFALAYRSIQNYISYLSVLCVSGGRKCSMAEIVGNDWE 567
QY 670 LMGDIPSVSTGLAYTIOHRNFKDKVHKHGFLELNYDGRFTNLISYAYQKSTQPTNF 729
Db 568 --GISDEYSDNMYIV-----NSASDVIAKGFLEMDYDAGFAFGRUSFSQOQTDQPTSI 620
QY 730 SDASESPNNASKEDQLKQGYGLRSVSLPRDYGRLEVGTRWLGKLTGLGAMRYFGKRSIR 789
Db 621 ASTY-----FGAGDMTELPKRYWTLDTGVRFDFNALTLGLTIKYTKARR 665
QY 790 ATABERYIDTNGTNSFNQLKGRSKSTQTTETLARQPLIFDFAAYEPKKNLIFRAEVKN 849
Db 666 LSPDFEQDEHTGA-----IIKQ--DLPIQIITIDLYGYEYRNLTKLSVQN 711
QY 850 LFDRIYDPLDAGNDAATQRYSSFPDKDDEVTGNADKTLGKNGYGGTSKSVLTNFA 909
Db 712 LMNDYSEALKNL-----MMPGLGDETHPANS-----AR 741
QY 910 GRTFLITMSYKF 921
Db 742 GRTWIFGDIRF 753

RESULT 6
Q92NX4 PRELIMINARY; PRT; 885 AA.
AC Q92NX4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE PUTATIVE IRON/HEME TRANSPORT PROTEIN.
GN SMC04205.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelie D., Purnelle B.,
RA Ramsperger U., Surzycki R., Thebaut P., Vandenbol M.,
RA Vorhoeelter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL Science 293:668-672(2001).
DR EMBL; AL591789; CAC46612.1; -.
KW Complete proteome.
SQ SEQUENCE 885 AA; 95389 MW; BEE36641060DFBF4 CRC64;

```

```

Query Match 9.6%; Score 468.5; DB 16; Length 885;
Best Local Similarity 22.8%; Pred. No. 1.2e-17;
Matches 228; Conservative 122; Mismatches 337; Indels 313; Gaps 42;

QY 14 LMGVTLHYSY-----AEDAGRGAQAQVLEDDVHVAKRVPKDK----- 55
Db 107 LLDGTGVHYQTANRAAVIGPAGDAGPGEAGATVLRIVVTGKTCRNANSAGAGCTP 166
QY 56 --VFTDARAYS-TRQDIFKS--SENLDNIVRSIPGAFQODKSSGI-VSLNIRGDSGFR 109
Db 167 DWYVEEPASVSVSDAVQASARNANDVLDVSAGVTSNRSEAGNPGIAINVRGLQDN 226

```

```

QY 110 VNTWVGITITFTYSTDAGRAGSSQFGASVDNSFNLAGLDVVKGSPSGSAGINSLAGSA 169
Db 227 VTTMIDGARQDFQ-----RAGHCASQRYVVDYTAFLRSVEVEKAVAGVAGGASLGGAV 279
QY 170 NLRTLGVDVVOGNTYKGLLKLTGTNSTKGNAMAAGARKWLESGASGVLYGHRRS 229
Db 280 NFRVTADDDIITPDGRGVNELNAETGTNAYFNG-SILGAARFSE---DPSVLGGISKR 335
QY 230 VAQNYRVGGGQHGIGNFAGBYLERRKORYFVQEGALKFNSDSGKWERDLORQOQKYPYK 289
Db 336 VG-DYDFGQNG-----KSPLLDLAVTTAVDDSDFLFSR----- 366
QY 290 NYNNOELQKY-----IEG-----HDKSWRENLAPOYDITPI-----DPSCLK 326
Db 367 -----LETFGLTKVEGSPDDFTFLSWLRN-----DSEATOGGLVFGDLRDDPONYL 415
QY 327 QQSAGNLFKLEYGCVFNKYTAQFRDLNFKIGSRKIIINRNYQFNGLSLNPYT-NLNLTAA 385
Db 416 NNTVSSSEFWDDPD-----SELIDKGLRWYNRVND-----ELROYTPPLPIYA 460
QY 386 YNSGRQKYPKSGFTGWLLKDFETYNNAKILDLNNTATFRLPRETELQTTGLFNY---F 442
Db 461 MTS-----FGG-----SLDNTSRF-----ETAL-GDLSLNYGGEA 489
QY 443 HNEYGKNRFPPEELGLFDFPDODNGL---YSYLGKRFKDGKGLLPQKSTIVQAGSYQNT 499
Db 490 YSDNGKTTTPPLV-----DQGFDEAYGKG-----LNPVGRSMTS 526
QY 500 FYFD-----AALKKDIYRLNYSNTVVG-----YRFGGEYTVGYGSDDEKRAFG 543
Db 527 AFLNATLEHDDWLEVGAGLRYDRYLRKAGFTEVGGRKPRYIVPGVCGYFYDDGEC---AYY 584
QY 544 ENSPTYKHCNRSCGIYEPVLKYGKRRANN---HSVISISADFGDYPMPFASYSRTHRMP 600
Db 585 DEDPVY-----GGGEAVLERVDIDKSGCALLPSARIAVMPPEGIQPFVYTAHTYRPP 636
QY 601 NIQEMYSQIGDSGVHTA-----LKPERRANTWQGFNTYKGLLKQDDTLGLKL 649
Db 637 SVMKAL-----TSGGHPGDAIATYIPNPKPGRGTWELGINIARDGLFTAGDSLRLKT 691
QY 650 VGYRSRIDNYI--HNIVYKQWMDLNGDIPSWSVSTGLAYTIQHRNFKDKVHKHGFLELNY 707
Db 692 VYFDRITODITLNGYFAFDKN-----LFQHVNLDDGDTMNGVEIEASY 737
QY 708 DYGRFTNLSYAKOSTQPTNFS-----DASESPNNASKEDQLKQGYLSRVSALPRDYGR 763
Db 738 DMGSAYVGASYTLKTDYADTYSYSGTASCTPLAASGNTVP-----SVLFVPPENKFT 792
QY 764 LEVGTWMLGNKLTGLGAMRYFGKSIRATABERYIDGTNGG--NTSNPRLGKRSIKOTET 821
Db 793 LDAGIRLFERKLVLGGRATY-----VSDSKPTVGQLAGLFTAGYK----- 833
QY 822 LARQPLIFDEYAAVEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFPDKDKE 881
Db 834 -----VEDIYGSYFSDSAKLRLAINNVDEQYAPALGA-----FYYP----- 871
QY 882 DVTGNADKTLGNGKYGCTSKSVLTNFAGRFTFLITMSYKF 921
Db 872 -----APGRATATVSLNFKF 885

RESULT 7
Q9X5P4 PRELIMINARY; PRT; 883 AA.
AC Q9X5P4;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HEME RECEPTOR HASR.
GN HASR.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.

```

```
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20121752; PubMed=10658665;
RA Ochser U.A., Johnson Z., Vasil M.L.;
RT "Genetics and regulation of two distinct haem-uptake systems, phu and
RL has, in Pseudomonas aeruginosa."
DR Microbiology 146:185-198(2000).
DR EMBL: AF127223; AAD31013.1;
DR InterPro: IPR005531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
KW Receptor.
SQ SEQUENCE 883 AA; 96942 MW; A29B07A3A6C31DCF CRC64;

Query Match 8.9%; Score 435.5; DB 2; Length 883;
Best Local Similarity 23.0%; Pred. No. 7.9e-16;
Matches 218; Conservative 117; Mismatches 375; Indels 239; Gaps 39;

QY 30 RAGSEAOIQVLEVDVHKAKRVPKDKKVFDDARAVST--RQDIFKS--SENLDNIVRSIPGA 86
Db 117 RVSQDLDVQMSPSV-ISAAR--PDDWVYQTPHSVSVIGREIERNPPRHAADMLEETPGV 173

QY 87 FTQODKSSGIVSNIRGDSGFRVNTWVGITOTFTYSTSTADAGAGSSQFGA-SVDSNF 145
Db 174 YSSVSQDDPGLSVNIRGIQDYGVRNMSVDGMQRNYQ-----QSGHQQRNGTLYVDPEL 226

QY 146 IAGLDVVKVSGSAGINSLAGSANLRTLGVDVVOGNTYG---LLLKGLTG-TNSTK- 200
Db 227 LSEVVIDKGASSAMGAGVIGGIANFRTLEARDLVPRGKVGVRVLTSGLGDANGTHF 286

QY 201 -GNAMAAIGARKWLESAGSVGLYGHSSRVAQNYRVGGGGQHGIFGAELERRRQRYF 259
Db 287 IGSAFAAIGTEVW-----DMLVAASERHLG-DYDPGTKGS-IGE----- 323

QY 260 VQEGALKFNDSGKWERDLQROQWKYP--YKNYNQELQKYIEGHDKSWRENLAPQYDI 317
Db 324 LRTGAW-FNPEAG-----QVRKHSVAYSGY-----VMSRLAKLGA 360

QY 318 TPIDPSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNGLSL-NP 376
Db 361 LPDQO---RLQFSYLTQVSYDDA-NMLTENQALMEKLGSSDVRAQNFADYGYAPDNP 416

QY 377 YTNLNLTAAY--NSGRQKYPKSGKFTGWGLLKDFEYNNNAKILDLNNTATFRPRETELQ 434
Db 417 LVDFKAKLYVDNRNRQOTLQRTIPGYSITYQTDYVG---AAQNTSTFALDDLTLR 472

QY 435 TTIGFNVFHNKYKNR-----FPEELGLFDDGPDQDNGLYSLGRFRKDGKGLLP 483
Db 473 ANYGLEFFYDKVRPDSQSPRASTSAVGFPAAEGW---TPKDRALGSLFARLDYD----- 524

QY 484 QKSTIVQAGSQYFNFTYFDAALKKDIYRL-----NYSNTVYGRFGGEYTYGSDDE 537
Db 525 -----YDDMLNLNAGLRYDRYLRGTGTGFNARTFLLGTTRQTDMPLOYAVD-- 570

QY 538 FKRAFGENSPTYYKHKHCRSGIGIEPVLLKYGKRRANNHVSISADFG-DYDFMFPASYSRT 596
Db 571 --REGRFSPTF-----GLSVKPGVDWLQFYATYKG 600

QY 597 HRNPNTQMYFSIGDSG-----VHTALKPERANTWQGFNTYKGLLKQDDTLGLKLV 650
Db 601 WRPPAYTESLITGRPHGGGAENMYPNPLSPERSKAWGVNFKENLWFSDDLRLGLKVA 660

QY 651 GYBSRIDNKHNYGKWDNLNGDIPSWSVSTGLAYTIQHRNFKDKYKHGFELELYNDYG 710
Db 661 YFTRVDDDFIFMGMGQO-----PPGYGMAGIGNSAYVNNL--DSTRFRGEVYQLDYDAG 712

QY 711 RFTNLISYAYQKS-----TQPTNFSDASESPNNASKEDQLQKQGYGLSRV-- 754
Db 713 LAYGQLSYTHMIGNSDFCSKTAWLGGVTVTKGSGRRPPVIMDRPDEQANATHCSAVLG 772

QY 755 --SALPRDYGRLEVGTRVLGNKLTILGAMRYFGKSTRATAERYIDGTNGGNTSNFRQLG 812
```

```
Db 773 SAEHMPMDRG-----SLTLG--MRFFORRLDVGARARYSEGYSVAGGATVSQAG 819
QY 813 KRSIKQTEITLARQPLIFDFAAYEPKKNLIFRAEVKNLFDRRYIDPLDAGNDAATQRYYS 872
Db 820 VYPADWKY-----TVYDLYGYSRVSDLEFLRLAMENVTRAYLVPLG----- 862
QY 873 SFDPKRDEDDVTNADKTLNCGKYGKTSKSVLTNFARGRTFLITMSYKF 921
Db 863 -----DV-----LAFTLGRGRTLOGTLEYQF 883

RESULT 8
Q9HYJ7 PRELIMINARY; PRT; 891 AA.
AC Q9HYJ7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HEME ACQUISITION PROTEIN HASR.
GN HASR OR PA3408.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL: AF004762; AAG06796.1;
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 891 AA; 97864 MW; 126D6F41E7D66979 CRC64;

Query Match 8.9%; Score 435.5; DB 16; Length 891;
Best Local Similarity 23.0%; Pred. No. 8e-16;
Matches 218; Conservative 117; Mismatches 375; Indels 239; Gaps 39;

QY 30 RAGSEAOIQVLEVDVHKAKRVPKDKKVFDDARAVST--RQDIFKS--SENLDNIVRSIPGA 86
Db 125 RVSQDLDVQMSPSV-ISAAR--PDDWVYQTPHSVSVIGREIERNPPRHAADMLEETPGV 181

QY 87 FTQODKSSGIVSNIRGDSGFRVNTWVGITOTFTYSTSTADAGAGSSQFGA-SVDSNF 145
Db 182 YSSVSQDDPGLSVNIRGIQDYGVRNMSVDGMQRNYQ-----QSGHQQRNGTLYVDPEL 234

QY 146 IAGLDVVKVSGSAGINSLAGSANLRTLGVDVVOGNTYG---LLLKGLTG-TNSTK- 200
Db 235 LSEVVIDKGASSAMGAGVIGGIANFRTLEARDLVPRGKVGVRVLTSGLGDANGTHF 294

QY 201 -GNAMAAIGARKWLESAGSVGLYGHSSRVAQNYRVGGGGQHGIFGAELERRRQRYF 259
Db 295 IGSAFAAIGTEVW-----DMLVAASERHLG-DYDPGTKGS-IGE----- 331

QY 260 VQEGALKFNDSGKWERDLQROQWKYP--YKNYNQELQKYIEGHDKSWRENLAPQYDI 317
Db 332 LRTGAW-FNPEAG-----QVRKHSVAYSGY-----VMSRLAKLGA 368

QY 318 TPIDPSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNGLSL-NP 376
Db 369 LPDQO---RLQFSYLTQVSYDDA-NMLTENQALMEKLGSSDVRAQNFADYGYAPDNP 424
```


QY 377 YTNLNLTAAY--NSGROKYPKSGKFTGWGLLKDPETNNNAKILDLNNATATRLPRETELQ 434
 Db 425 LVDFKAKLYVDNENRQOTLQRTGTPGYSITYQTDYTG----AQAOQNTSTFALDDLSTLR 480
 QY 435 TTLGFNFHNEYGNR-----FPEELGLFFDGPDDNGLYSLYGRFGKDKLLP 483
 Db 481 ANYGLEFFYDKVRDPSQPRASTSAVGFPAAGM---TPKGDRAIGSLFARLDYD----- 532
 QY 484 OKSTIVOPAGSOYFNTFYFDAALKKDIYRL-----NYSTNTVYGRFGGYTYGYSDDDE 537
 Db 533 -----YDDLNLNAGRLRYDRYLRGDTGFNARTFILTTRQTDMPLOYAYD-- 578
 QY 538 FKRAFGENSPYKHKCNRCGIEYVPLKYYKKRANHNSVISADFG-DYFMPFASYSRT 596
 Db 579 --REEGFRSPTF-----GLSVKPGVDMLQLFATYK 608
 QY 597 HRMNIQEMTFSGDSG-----VHTALKPERANTWQFGFNTYKGLLQDDDTFLGLKV 650
 Db 609 WRPPAVTESLITGRPHGGGAENMYPNFPLSPERSKAWEVGFNVLKENLWFSDDDELGLKVA 668
 QY 651 GYRSRIDNYHNHYGKMWDLNDIPSWSTGLAYTTOHRNFKDKVHKHGFELNELNYDG 710
 Db 669 YFDTRVDDEFPMGMQ-----PPCYGMAGICNSAYVNNL-DSTRFRGVVEYQLDYDAG 720
 QY 711 RFTNLISYAYOKS-----TOPTNFSDAESPNNASKEDQLKOGYGLSRV-- 754
 Db 721 LAYQLSYTHMIGNSDFCSKTAMLGVTQTVKSGRRPPVIDMRPEQANAAHCHSAVLG 780
 QY 755 --SALPRDYGRLEVTGRLGNKLTGLGAMRYFGKISIRATAEERYIDGTNGNTSNFRLG 812
 Db 781 SAERHPNDRG-----SLTLG--MREFDRLDVGARARYSEGSYVAGGATVSOAG 827
 QY 813 KRSTKQETLARQPLIFDFAAEPEKKNLIFRAEVKNLFDORRYIDPLDAGNDAATQRYYS 872
 Db 828 VIPADWKEY-----TVYDLYGSYRVSDELTLRLAMENVTDRAYLVPLG----- 870
 QY 873 SFDPKDKDEDVTCNADKTLGCKYGGTSKVLNTNFARGRTFLITMSYKF 921
 Db 871 -----DV-----LAFTLGRGRTLQGTLEYQF 891

RESULT 9
 Q9HUR6
 ID Q9HUR6 PRELIMINARY; PRT; 989 AA.
 AC Q9HUR6
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE HYPOTHETICAL PROTEIN PA4897.
 GN PA4897
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey C.K., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL; AE004902; AAC08282.1; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 989 AA; 108303 MW; 99D332FBD0F0806F CRC64;
 SQ

Query Match 8.8%; Score 428.5; DB 16; Length 989;
 Best Local Similarity 23.0%; Pred. No. 2.2e-15;
 Matches 229; Conservative 132; Mismatches 375; Indels 261; Gaps 42;

QY 30 RASEAQIOVLEDDVHVAKRVPKDKVFTDARAVSTQDIFKSSENLD-----NIVR 81
 Db 149 RMSGEAPADLSPPVVSAAELADPOKETYT-----APRSSVYLSEEDIRGRVSVGDLQ 203
 QY 82 SIPCAFTQODKSSGIVSINIRGDSGFRVNTVMDGTQTFTYSTDAGRAGGSGQFASV 141
 Db 204 GIPCVQVDSRNGALDVNIRGIQOSRVAVRVDCAEQ-----ALDVYRGVAGTQORSYI 258
 QY 142 DSNFIAGLDVYVGSFSGSAGINSAGSANLRLTGLVDDVQGNNTYGLLLKGLTGTNSTKG 201
 Db 259 DPDLVSVTVTKGSPSTRGAI---GGSVEMRTIGVKDILVDGKDLGVFRFTGDVNN--G 312
 QY 202 NAMAAGARKWLESASV---GVLYGHSRR---SVAQNYRVGGGQHIGNFGRAEYLERR 254
 Db 313 VAPQHSASSKTENLSSVPHDDRGSLFSGSAKSAFAFAYR---NEHL-DLVAAYAQNR 367
 QY 255 KORYFVQEGALKFNSDSGKWERDLQROQWKYPKYNNNOE---LQKYIEGHD----- 304
 Db 368 QGNYF-----SGKKQD-----RYRVNRYGRESSVAKYVYVNAEEVLSNSE 410
 QY 305 -----KSWRENLAPOYDIT-----PIDPSLKQOSAGNLFKLEYDGV-FNKY 345
 Db 411 TESYLKATWR--LADETLGLYRRYDGRGTGEIMPSDIFRGTAGIYQVPLSEVKIDTY 468
 QY 346 TAQERDLNTKIGSRKIIINRYQFNYSGLNAPYTNLNLTAAYNSGRQYKPGSKETGWGLL 405
 Db 469 TARYRYPEN-----NPLVDLSTGLWMTAKSMDLTSVLAPRSQAYRSRNNWT----- 516
 QY 406 KDFETYNNAKI-LDLNNTATFRL-----PRETELOTTLGFN----- 440
 Db 517 ----RQDNRRIIGDLNNVARETDFGDKLDGGSFOVEDIQPKSVVTVTLHDINANRTL 572
 QY 441 --YFHNIEYGN-----RPPEELGLFFDG-----PDQONGLYSLGRFKGDKGLLPKSTI 488
 Db 573 RDATRQYGLNGKLEFRPVERLTLMGGGRYSHFNSKNGISASPRREDRMRFT---TV 628
 QY 489 VQPA--GSOYF---NTFYFDA--LKKDIYRLNYSNTVYRFGGEYTYGYGSDDEBK 539
 Db 629 SRPGYYSMMWFFPDQNGQYTDATDPRLNGGIVTNTNPPFEGIF-----DEFG 677
 QY 540 RAFGENSPYKHKCNRCGIEYV-LKKYKGRANNHNSVISADF---GDYFMPFASYR 595
 Db 678 PA-----NVTVHPSRVTNVVTVGYNSKKGSSRGGSFAPFGINFELAPDFTV-YASYTE 730
 QY 596 THRPNTQEMYFSGISGVHTA--LAPERANTWQFGFNTYKGLLQDDTLGLKLVGR 653
 Db 731 GLRLPSLFE---TSQGTLOVEPGKDLKPEPERSRWEIGASALRDSLLADGDSAAIKLAYFN 787
 QY 654 SRIDNYHNHYGKMWDLNDIPSWSTGLAYTTOHRNFKDKVHKHGFELNELNDYGRFF 713
 Db 788 NTKANYTRYD-----PGQGLMTFSNT-----DSYRTSGLELQSHDAGRVF 831
 QY 714 TNLISAYOKSTQPTNFSDAESPNNASKEDQLKOGYGLSRVSAL-----PRDYGR 764
 Db 832 ADLSATYLYLTETCDAAFAARLAGANRYQRTENTPCTPGSPMGSYTNTQNPRLATNL 891
 QY 765 EVGTRWLGKNTLGGAMRYFGKSTRATAERYIDGTNGNTSNFRLGKRIKOTETLAR 824
 Db 892 TAGLRFEDQALTLGRMTY-TSGGTATADKPKWQV-----ATTPQIEYRSVQ----- 937
 QY 825 QPLIFDYAAEYPAKKNLIFRAEVKNLFDORRYIDPLDAGNDAATQRYYSFDPKDKDEDVT 884
 Db 938 ---LFDLFLYKLFKEHTELNLSQNLTDRYLDPL-----AQSFMP----- 975
 QY 885 CNADKTLGCKYGGTSKSVLTNFARGRTFLITMSYKF 921
 Db 976 -----APGRTLVRGMQAKF 989

```
RESULT 10
Q9RHT4 PRELIMINARY; PRT; 916 AA.
AC Q9RHT4; 2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE HASA RECEPTOR PROTEIN.
GN HASR.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NO.33;
RX EMBL; AB023289; BAA88490.1;
RA Idei A., Kawai E., Akatsuka H., Omori K.;
RT "Cloning and characterization of the Pseudomonas fluorescens ATP-
binding cassette exporter, HasDEF, for the heme acquisition protein
HasA."
RL J. Bacteriol. 181:7545-7551(1999).
DR EMBL; AB023289; BAA88490.1;
DR InterPro; IPR001589; Actinin_act_bind.
DR Pfam; PF00593; TonB_boxC.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 916 AA; 101187 MW; CF392CAA8939EDC8 CRC64;

Query Match 8.7%; Score 426; DB 2; Length 916;
Best Local Similarity 22.4%; Pred. No. 2.7e-15;
Matches 226; Conservative 129; Mismatches 372; Indels 276; Gaps 45;

QY 22 YSAEDAGRAGSEAIQ--VLEDVHVAKRVPKDK-----VFT 58
DB 87 YSAEGLGLVGSAPVEYRFGDGRQVLTTRVSADHGDALTVGTTITGNTGKPSDWYD 146
QY 59 DAR--AVSTRQDIFK--SENLDNIVRIPGAFQODKSSGIVSLNIRGDSGFGRVNTWVD 115
DB 147 TPRSVAVISREIQIDKRPRIADMLEETAGVYTAVNQRDPLSLNIRGVQDYGRVNNID 206
QY 116 GITQFYSTDADGRAGSSQFGAS--VDSNFIAGLDVVKSGSPGSAGINSLAGSANLRT 173
DB 207 GMQNF-----NVNGHQORNGTMLIDPEFISIEIDKSGSGQGGAVALGSIASFKT 258
QY 174 LGVDVVYQGNNTYGLLLKGLTGT----NST--KGNMAAICARKWLESGASGVLYGHSR 227
DB 259 LEASEFLADGKDYGRLRAGSGIGELNGTYFNGSGVFAFGD---ERG---DVLLGYSE 311
QY 228 RSVANQYRVG--GGQHGHTGNFGAEYLERRKORYFVQEGALK--FNDSGKWERDLQROQWK 284
DB 312 RHFG--DYRAGTHNDQKLTG----HLRADS-----QPAAFDDDLNLSSEVGTGSTRSIVK 362
QY 285 Y---KPKYNNQBLQ--KYIEGHDKS---WR-----ENLAPQYDITPIDPSLKQOQSAGNL 333
DB 363 FGLNLP-----NDQRVOLSYLESDDSDNDANAYTAPDNQSVYQ--RVSKNNLNKAKNVGLD 416
QY 334 FKLEYDGVFNKRYTAQFRDLNKTGSRKIINRNYQFNVGLSLNPTNLNLPAAVNSGRQKY 393
DB 417 YRYTPDNPMDIFRAKVYVVTQMDRENAPN-----AATLTSNGYVAGVTDHFQT- 465
QY 394 PKGSKFTGWLGLDGFENYNAKILDLNNTATFRLPRETELQTLTGFNYFNEY--GKNRF 451
DB 466 -----DTWQVQGD-----NTRFDFDGLGHVSRNYGVVEYQDKFKPSTNKV 506
QY 452 PEELGLFFDQDQDNGLYSL--GRFGKDKGLLPQKSTIVQPAQSQVFNTFYFAALKKOI 510
DB 507 -EATNL-----SGLLPVEVGGTGGKRTI---ASLFNNLQYEGDWLTLEAGLYDR 554
```

```
QY 511 YRLNYSTNTVGYRFGGEYTYGSDD-----EFKRAFGENSPYTKKHCNRSCTGIPEVLK 565
DB 555 YRLEGVTGMLTYRRDRFYSSTVGAKRVVEEVDIDREEGQFSPTF-----GIGI-KPCL- 606
QY 566 KYCKKRANHSVISASADFGDYFMPFASYSRTHPNQEMVFSQIGDSGVHTA----- 618
DB 607 -----DWMQLYARWGKWRPPAVTETFTMTGRPHGSSSERVFPNPF 647
QY 619 LKPERANTWQGFNTYKGLLKODDTLGLKLVGYSRIDNYI---HNVYCKWMDLNGDIP 675
DB 648 LKPEESRDWEVGVNFKEGLLFGDGRGMKVAYEDTRIENFSLNHSV-----SLP 698
QY 676 SWTSSTGLAYTIQHRNFKDKVHKHGFELNLDYDGRFETNLSA----- 719
DB 699 E-TSVGGFLGTMAVYVNTNTRFRGVEYQLNYDMGRAYANLSYTHMIGSNEFCSKNYMG 757
QY 720 -----YOKSTOP-----TNFSDASESPNNASKEDQLKOGYGLSRVSALPR 759
DB 758 GAKKNGPSTTRYERYTRPNTGTIGLRPVTVTVTEVLDDDAAN-NKE---SCGRIMGNATYMPA 813
QY 760 DYGRLEVGTEWLGKLTGLGAMRYFGKSIRATAEERYIDGTNGGNTSNFRLGKRISIKOT 819
DB 814 DRGSLTLGARFLKRLKLDGMGVRY-----SSNGENLDSQGYDFMDQ- 855
QY 820 ETLARQPLIFEDFYAAAYEPKKNLIFRAEVKNLFRDRIIDPL-DAGNDAATQRYYSFDPKD 878
DB 856 -ALWPQYTLVLYASYWMTDQLNIALALENATDEAYFVAMGDANN----- 899
QY 879 KDEDVTCNADKTLGNGKYGTSKSVLTNFARGRTFLITMSYKF 921
DB 900 -----LSLARGRTLGMLEYKF 916

RESULT 11
P72473 PRELIMINARY; PRT; 899 AA.
AC P72473;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HASR PROTEIN.
GN HASR.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SM365;
RX MEDLINE=97315228; PubMed=9171402;
RA Ghigo J.M., Letoffe S., Wandersman C.;
RT "A new type of hemophore-dependent heme acquisition system of Serratia
marcescens reconstituted in Escherichia coll."
RL J. Bacteriol. 179:3572-3579(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SM365;
RA Binet R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y08983; CAA70172.1;
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
SQ SEQUENCE 899 AA; 98281 MW; 3A13AE4CFCA3911D CRC64;

Query Match 8.4%; Score 412; DB 2; Length 899;
Best Local Similarity 23.4%; Pred. No. 1.6e-14;
Matches 221; Conservative 129; Mismatches 323; Indels 270; Gaps 48;

QY 56 VFTDARAVSTRQDIFKSSSENLDN-----IVRSIPGAFQODKSSGIVSLNIRGDSGF 107
DB 150 VYDEPRSVSV-----ISREQMDNRPARHADILEQTTIGAYSSVQSDPALSVNIRGIQDY 204
```

QY 108 GRVNTWDTGTTQTYSTSTADAGRAGGSSQFCASVDSNFIAGLDVVKGSFSGSAGINSLAG 167
Db 205 GRVNNIDGMQRNFKSG--HGQRNGTMY-----IDSELLSGVTIDKGTGGMGSGAGTLGG 258
QY 168 SANRLTLGVDDVQGNNTYIGLLKGLTGNTSK--GNAMAAIGARKWLESASGVVLYGH 225
Db 259 IATNTVYSASDFLAPKELGKLHASTGDNTHIGSGILALGN---ETG---DILLAA 311
QY 226 SRRSVAQNYRVGGGQHIGNFGABYLERRKORYFVQEGALKFNDSKQWDLRQOQWKY 285
Db 312 SERHLG-DYMPGNKGD-IGN-----IRINDTGNDRYAESIKNNK 350
QY 286 KPYKNYN-----NOELO-KYIEGHDKSHRENLAPOYDITPIDPSLKQOSA 330
Db 351 IPDTHYRMHSLAKVGNLPAQRQLSYLQTQTAS-----PI-----A 389
QY 331 CNLFKLEYDGVFNKYTAQFDRLNTKIGSRKIINRNYOFNYSLSNPYTNLNLTA-AVNSG 389
Db 390 GTLINL---GTRPPYELGWK---RTGYDVMARNAADYSLAPEDVDWLDFOAKLYVD 442
QY 390 RQYPKGSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRE--TELQTTLGFNY-----F 442
Db 443 TQ-----DDSDTYSTSSLLD-NGYATRLRITYGAQAQNTSRFSLAPGHDF 487
QY 443 HNEYGNRFEPELGLFFDQDODNGLYSYLGKFKDGKGLLPQKSTIYQVAGSQYENTFY- 501
Db 488 RANYG-----LEFYDKAFSSSR-----QCMGVTGAGNRV---ASLFANLTYD 530
QY 502 -----FDAALKDILYRLNYSNTVGY-----RFGGEXTGYGSDDEK--RAFGENSPY 549
Db 531 YDGLWLTBGLGLRYDRYLRGTG--LSYPLDAKDGORYT----IDNPKALRLTCCSTTR 585
QY 550 KK-HCNRSCTGIYEPVKKYKGRANNHVSISADFGFMPFASYSRTHRPNQIOMYFS 608
Db 586 EDWDVDRDQGLSPTL-----AVAVRPGV-EMLELYTYGKSWRPPAITE----- 629
QY 609 QIGDSGVHTA-----LKPERANTWQGFNTYKGLLKODDTLGLKLVGYRSIDNYI 660
Db 630 TLTGSAHSSSTQPNPLQPERGAMVEGVNQVQPDWLFEGDLVAKVGYDFKVDNYI 689
QY 661 HNVYKGMWDLNGDIPSWYS-STGLAYTIQHNRNFKDKVHKHGFELNLYDYGRRFTNLSYA 719
Db 690 NLA-----IDRNKPLGVQPSIGNAAVY---NLSKTRFGLYQLNVDAGVFADLTYYT 740
QY 720 YQ--KSTOPTN-----FSDASEPN-----NASKEDQKQGYGLSRVSA--LPRDY 761
Db 741 HMIKNEFCNSKAWLGLRLYGDGSRGRNFYVEPDAAASDNFVTCDDGTFQFGSAAAYLPGR 800
QY 762 GRLEVTGRLWGLKTLGGAMRYFGKSIIRATAERYIDCTNGNT-SNPRQLGKRSIKOTE 820
Db 801 G-----SVTLGG--RAFORKLDAGVTVRFAPGYQDSSVPSNY-----P 836
QY 821 TLARQP--LIFDYAAYPKKNLIFRAEVKNLFDORRYIDPDAGNDAATQRYSSFPDK 878
Db 837 YLADWPXYTLDELVASYKLTDSLTLRGSVENLTNRAVY-----874
QY 879 KDEVTGNADKTLONGYGGTGSKSVLTFNFRGRFTLITMSYKF 921
Db 875 -----VSYGETLANTL-----GRGRTVQGGVEYRF 899

RESULT 12
Q9CKJ4
ID Q9CKJ4 PRELIMINARY; PRT; 848 AA.
AC Q9CKJ4;
DT 01-JUN-2001 (TremBrel. 17, Created)
DT 01-JUN-2001 (TremBrel. 17, Last sequence update)
DT 01-DEC-2001 (TremBrel. 19, Last annotation update)
DE HASR.
GN HASR OR PM1622.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.

NCBI_TaxID=747;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=PM70;
RC MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR EMBL; AE006199; AAK03706.1;
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 848 AA; 95909 MW; 5C6B28E913FD583 CRC64;

Query Match 7.3%; Score 357.5; DB 16; Length 848;
Best Local Similarity 21.6%; Pred. No. 1.4e-11;
Matches 207; Conservative 153; Mismatches 364; Indels 235; Gaps 49;

QY 12 FYLMGVTLHYHYSYADAGRAGEAIOVLEDVHVAKRVPKDK-----KVFTDARAYS- 64
Db 14 FLMLGLPTWAFSQAN-----LEKSTINKLETILVNESE-EKNKFDENLIKTYLSSGSYSY 67
QY 65 -TRODI--FKSSENLDNIVRSIPGAFTQODKSSGIVSLNIRGDSGFGRVNTWVDGITQTF 121
Db 68 LSQSDISTFRGS-SVGDFLGPVGVIVGNKRNKSGALSNIIRGIANENRVPVVID---KGL 123
QY 122 YSTSTADAGRAGGSSQFCASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVYQ 181
Db 124 QSVPSYQYAGSSYR--TYLDPDLISQVEIEKSPSLMDATGATGCVVYVETLRWQDIIP 181
QY 182 GNNTYGLLLKGLTGNS-----TKGNAMAAIGARKWLESASGVVLYGHs-----RRSV 230
Db 182 QGKNWGVRLKGLTMTNTVSPPPYTRGGYQT-----KYISK-----CLSNHTGLCQTQY 231
QY 231 AONRVGCGGQHIGNFG-----AEYLERKORYFV-ORG-----ALAFN 268
Db 232 APNARYSHGFDLNAYNYSLAFANKWNADLVAYAKRKQGNFVGRHGQTPVIESIEFE 291
QY 269 SDSGKWERDLQROWKYKPYKNYNQOELQYIEGHDKSWRENLAPOYDITPIDPS-SLKQ 327
Db 292 EDS-----VEVKEPRVHEVEIGSLTFKENRSTLY--RGEALNLSQ 332
QY 328 QSAGNLKLE-YDGY-----FNKYTAQFDRLNTKI-----GSRKIINRNTYOF 368
Db 333 DNTSYLAKINYNVDHRLGLAYRHYHSRFGEMSSILNFRAYGALQEGTEVKVD-SYHA 391
QY 369 NYGLS-LNPYTNLNLTAAYNSGRQKYPKGSKFT-----GHGLLKDFETY--NNAKILDL 419
Db 392 NYSYNPTTPYVNLNVAIFYTDS-----DSSNFTPIEYGYSLSRHAHFLVSKOKGLSI 446
QY 420 NNTATFRL-PRETELQTTLGFNHYNEYGKNR-----PPEELGLFFDQDODNGLYS 470
Db 447 ETSIFQLNDKPFSLKYGLAHSYERIYOPNAQARVAKGYPEDA----IGP-----LYI 497
QY 471 YLGRFKGDKGLL-----POKSTIVQAGSQYENTFYDAAKKDIYRLNYSNTNTVGYRFG 526
Db 498 RDGKRKEWSAFVAANYPITSMLKADIGLYLQSTIYD-----YIVRTERVNIIG 546
QY 527 EYTGYYGSD-----DEFKRAFGENSPYKHKCNKSCGI-YEPVLKYYKKKRNHSHVSIS 580
Db 547 AFVNPNGPGNIWVEKYKDYVVKQAPVKNKGMSPVMTLTFEPI-----NGVQI- 594
QY 581 ADFGDYFMPFASYSRTHRPNQIOMYFSQIGDSGVHTA-----LKPERANTWQGFENTYK 635
Db 595 -----YTKYAEALRSPS---LFQATKGSWSATADNLEQLPERAKKWEAGINLIFY 642
QY 636 KGLLKQDDTLGLKLVGYSRSDNYTHNYYGKWDNLGDIPIWSVSTGLAYTIQHNRNFKDK 695
Db 643 ENLGGKNILGFLKAYFNRIKDYLTRSYSP-----KDKVTQT-----INQSAQPK-- 689
QY 696 VHKHGFELNLYDYGRFTNLSYAYQKSTQPTNFSDESPPNASKEDQLKQGY-----GL 751

```
Db 690 -----GIELSAIDYDMKFKYAKLAGTYYTKT-----ECLTAEE-----QAGKEQCNSGYIYRNSL 739
QY 752 SRVSALPDYGRLEVGRLWGLNKLTLGCAMRYFKS-IRATAERYIDGTNGGNTSNFRQ 810
Db 740 NN-AVPPRLNLHATGLRFLFKQKLDIGARVSYYSKRILVPLVLSAERFV-----NTSIE- 791
QY 811 LGRSIRKOTETLARQPLIFDYAAEYKPKNLIFRAEYKVNLFDRRYIDPLDAGNDAATOR 869
Db 792 -----WAPYSLV-DLIYANYNSNKLKLTWTDWNVFNYRLDINNGLNTAPGR 838

RESULT 13
Q98L70 PRELIMINARY; PRT; 747 AA.
AC Q98L70:
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE HEME ACQUISITION PROTEIN, HASR.
OS MLR1155.
GN Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002996; BAB48593.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 747 AA; 80102 MW; 4105DA001B73E752 CRC64;
```

```
Query Match 7.1%; Score 346; DB 16; Length 747;
Best Local Similarity 21.4%; Pred. No. 4.9e-11;
Matches 197; Conservative 120; Mismatches 303; Indels 302; Gaps 38;

QY 23 SYAEDAGRAGSAQIOVLEVDVHKAKRVPKDK-----KVFTDARAVSTRQDIFKS 72
Db 52 AHAQAQTAPAGEQTDQ-----SKADOEKAATAGATLLDKILVLSRTGETAIESLAS 103
QY 73 SENLD-----NIVRSIPGAFTQDQKSSGIVSLNIRGDSFGGRVNTMVDGITO 119
Db 104 ASHVDQQLAHRMATTPNEMLLGPVGATQADARRVSTSNIRGLQDFGRVAVIVGARQ 163
QY 120 TFFYSTSTADAGGSGOFGASVNSFNITAGLDVVKVGFSGSAGINSLAGSANRLTLGVDDV 179
Db 164 DFO-----RSDHGTOSTYIDPELVKSVDIRGVPVANTYGSAGVGVFFEDTKDAADF 216
QY 180 VQGNNTYGLLLKGLTGNSTNKGMAAIGARKWLESASVG-VLYGHSRVSQAQNRVGG 238
Db 217 LKPEETWGSVTVGYESNG-KGWTTSATGATGYREFNENWDALGNIVYRN-----YDNYK-DG 269
QY 239 GGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSKWERDLQROWKYPKYNNOELQK 298
Db 270 GGVTVNGTGDVL-----SGLLKTS-----IRPTENSELKL 300
QY 299 YIEGHDKSWRENLAQYDIPDPSSLKQOSAGNLFKLEYDGVFNKYTAQFR----- 350
Db 301 GWGSSDGDWDETSGM-----PVNDVCLKS-----NFTARYNTIDEDKSW 341
QY 351 -DLNTKIGSRKIIRNNYQFVGLSLNPNYTNLNTAAYNSGR-----OKYPKGSKFTG 401
```

```
Db 342 LDLH-----INTSYN-----KTNLDLTSLVPQKRFDPITGLPTVLPAQSQT- 383
QY 402 WLLKDKETNNAKILDNNNTATFRLPRETELQTTLGFNFHNEYKRNRPPEELGLPFDG 461
Db 384 -----PDVGTTC--IDIWNTSRF-----ETGGTAHELTYGGDVGWV 416
QY 462 PDQDNGLYSVLGRFKGDKGLLPQKSTIVQAGSQYENFVFD-----AALKKDIYRLNYS 516
Db 417 DVKTG-----GTAGGDSFTYPSGKRNW--SGATVQDKLTWDLVLEIAGLRYDYSKDS 469
QY 517 T-NTVGYRFGEXYT-GYVGSDEDFKRAFGENSPYKKHCNRSCGIYEPVLKKYKKRANN 574
Db 470 TRETSGDRSLSPRITGV-----SPF-----ESAGL--AGLQFYG----- 501
QY 575 HVSISADFGDYFMPFASYSTRHMPNIQMYFSQIGDSGV-----HTALKKPERANTWQ 628
Db 502 -----TYAEGYRSPSLTETLISGNHAGVTFPFPFLPNLPNRPETGKTTE 544
QY 629 FGSNTYKKGLLKODDTLGLKLVGRSRIDNVYHNVYKGMWDLNGDIPSWVSSTGLAVTIQ 688
Db 545 FGINYRQNDIFEPGDALRVKAAVFNHNNVDYIDCVTLSPFAPGSGCP---FGPQIPICFQ 601
QY 689 HRNFKDKVHKHGFELNLYDYGRFNTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQG 748
Db 602 YQNFQ-AQKIDGFELEGVYDAGWYAGLSAS-----ITNG 635
QY 749 YGLS-----RVSNALPRDYGRLEVGTEWLNKLTGLGAMBYFGKSIKIRATAERYIDGTNG 802
Db 636 HTISYKGEADLATIPSSQVTAQLGRLEDKLTGVGEVYNGKP-----K 681
QY 803 GNTSNFRQLGKRSIKOTETLARQPLIFDFVAAVEPKKNL--IFRAEYKVNLFDRRYIDPLD 860
Db 682 GN-----AVAEDYTLVNAFASQATNDLKVDFRAD--NLFDVKYANPLN 723
QY 861 AGND-----AATORY 870
Db 724 GSTTAVVYEPGITLKLAAATMRF 745

RESULT 14
Q93SH4 PRELIMINARY; PRT; 782 AA.
AC Q93SH4:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HMUR PROTEIN.
GN HMUR.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1105PC4;
RA Nienaber A., Hennecke H., Fischer H.M.;
RT "Discovery of a haem uptake system in the soil bacterium
RT Bradyrhizobium japonicum.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ311165; CAC38746.1; -.
SQ SEQUENCE 782 AA; 83713 MW; F8FC1EBFE568AA87 CRC64;
```

```
Query Match 6.4%; Score 311; DB 2; Length 782;
Best Local Similarity 20.9%; Pred. No. 4.3e-09;
Matches 192; Conservative 99; Mismatches 317; Indels 310; Gaps 39;

QY 30 RAGSEAOQVLEVDVHKAKRVPKDKKVFVDARA----VSTRQDIFKSSNLDNIVRSIPG 85
Db 71 QAGSAAPVOTLDTITVAA---TKTRERADALAPVSSISLDQIQGLQPNKLSDFVHSVP 127
QY 86 AFTQDQKSSGIVSLNIRGDSFGGRVNTMVDGTTQTFYSTSTADAGRAGGSSQFGASVDSNF 145
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:44:38 ; Search time 62.3 Seconds
(without alignments)
1643.821 Million cell updates/sec

Title: US-09-762-926-4
Perfect score: 4904
Sequence: 1 MRSSFRLPFCPLMGVMLY.....SVLTNPARGRPLTMSYKF 922

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4904	100.0	922	21	AAV69381
2	4894	99.8	922	21	AAV69380
3	4779.5	97.5	921	21	AAV69382
4	4754	96.9	922	20	AAV38940
5	4743	96.7	922	20	AAV38939
6	4599.5	93.8	888	20	AAV38937
7	4413	90.0	871	20	AAV38938
8	2598	53.0	947	21	AAV44428
9	2514.5	51.3	913	21	AAV94671
10	2489	50.8	918	21	AAV94672
11	1976	40.3	393	20	AAV38936

12	900	18.4	753	22	ABB52934	Escherichia coli p
13	900	18.4	753	22	ABB52968	Escherichia coli p
14	278	5.7	915	20	AAV07476	N.gonorrhoeae trans
15	278	5.7	915	21	AAV51770	N. gonorrhoeae str
16	278	5.7	915	21	AAV80374	N. gonorrhoeae str
17	269.5	5.5	908	21	AAV51768	N. meningitidis str
18	269.5	5.5	908	21	AAV80372	N. meningitidis str
19	262.5	5.4	971	21	AAV21223	Haemophilus somnus
20	261	5.3	909	20	AAV07477	N.meningitidis tra
21	260	5.3	791	17	AAV95566	N. meningitidis se
22	259	5.3	790	17	AAV95565	N. meningitidis se
23	257	5.2	911	21	AAV51769	N. meningitidis str
24	257	5.2	911	21	AAV80373	N. meningitidis str
25	257	5.2	915	22	AAE12018	Neisseria meningit
26	255	5.2	908	15	AAV48220	N.meningitidis IM2
27	250	5.1	884	14	AAV34402	Sequence of low mo
28	245	5.0	912	16	AAV96969	Transferin recept
29	238	4.9	790	17	AAV95568	N. gonorrhoeae B h
30	237.5	4.8	792	17	AAV95567	N. meningitidis se
31	233	4.8	912	16	AAV77886	Bacterial transfer
32	233	4.8	912	18	AAW08959	Amino acid sequenc
33	233	4.8	912	18	AAW08961	Amino acid sequenc
34	233	4.8	912	19	AAV53044	H. influenzae str
35	233	4.8	912	19	AAV53046	H. influenzae str
36	233	4.8	912	21	AAV51689	H. influenzae type
37	233	4.8	912	21	AAV51691	H. influenzae type
38	233	4.8	912	21	AAV80355	H. influenzae type
39	233	4.8	912	21	AAV80357	H. influenzae type
40	232.5	4.7	913	16	AAV77884	Bacterial transfer
41	230.5	4.7	911	16	AAV77892	Bacterial transfer
42	230.5	4.7	911	18	AAW08966	Amino acid sequenc
43	230.5	4.7	911	19	AAW54124	H. influenzae str
44	230.5	4.7	911	21	AAV51780	H. influenzae non-
45	230.5	4.7	911	21	AAV80362	H. influenzae str

ALIGNMENTS

RESULT 1
AAV69381
ID AAV69381 standard; Protein; 922 AA.
AC AAV69381;
XX
XX
DT 19-JUN-2000 (first entry)
XX
XX
DE A BASB024 outer membrane protein of N. meningitidis.
XX
KW BASB024; outer membrane protein; N. meningitidis infection;
KW bacteremia; meningitis.
XX
OS Neisseria meningitidis.
XX
PN WO200011182-A1
PD 07-MAR-2000
XX
PF 13-AUG-1999; 99WO-EP05989.
XX
PR 18-AUG-1998; 98GB-0018004.
XX

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Thonhard J;

WPI; 2000-224702/19

N-PSDB; AAZ6158T

Novel polypeptides derived from the products of the BASB024 gene of Neisseria meningitidis, useful for inducing an immune response and producing antibodies useful for treating meningitis -

PS Claim 3; Page 88-91; 103pp; English.

XX The present sequence represents a BASB024 outer membrane protein of

CC Neisseria meningitidis. The BASB024 polynucleotide sequence was

CC first identified in the Incyte Pathoseq database containing

CC unfinished genomic DNA sequence of N. meningitidis. BASB024

CC polypeptides and polynucleotides are useful for generating an

CC immune response in an animal. Antibodies specific BASB024 polypeptides

CC are useful for treating N. meningitidis infection, which causes

CC bacteremia and meningitis.

XX

SQ Sequence 922 AA;

Query Match 100.0%; Score 4904; DB 21; Length 922;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 922; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSFRKLPICFYLMGVMLYHHSYAEADAGRAGSEAOQVLEDVHVHAKRVPKDKKVFDTDA 60

Db 1 mrssfrlkipcfylmgvmlhyhhsyaedagragseaqlvledvhvakrvpdkkvvftda 60

Qy 61 RAVSTRQDIFKSSNLDNIVRSIPCAFTQODKSSGIYSLNIRGDSGFGFVNTWVDGIGTQT 120

Db 61 ravstrqdifkssnldnivrsipgafatqgdkssgvlslnirgdsfgfgrvntmvdgigtq 120

Qy 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVYVKGFSGSGAGINSLAGSANLRTLCGVDDVV 180

Db 121 fystsdagragssqfgasvdsnfiaagldvvyvkgfsfgsaginslagsanlrtlcvddvv 180

Qy 181 QGNNTYGLLLKGLTGCTNSTKNMAAATGARKWLESAGSVGLYCHSRSSVAQNYRVGGGG 240

Db 181 qgnntyglllkgltgctnstkgnmaaatgarkwlesagsvglvchsrssvaqnyrvgggg 240

Qy 241 QHIGNFGAEYLERKQRYFVEGGLKFNSSGKWERFORPYWKTWKYQKYNDDPQELQKY 300

Db 241 qhignfgaeylerkqryfveggllkfnssgkwerdfqpywktwkyykynddpqelqky 300

Qy 301 IEHGDKSWRENLAFOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQFRDLNLTKIGSR 360

Db 301 ieghdkswrenlapoyditiPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQFRDLNLTkigsr 360

Qy 361 KIIRNRYQFNGLSINSVANLNLAAYNSGRQKPKSGKFTGWLGLKDFETYNNAKILDL 420

Db 361 kiirnrqfnvnglsinsvanlnlaaynsgrqkpksgkftgwlglkdfetynnakildl 420

Qy 421 NNTATFRPRETELOTTILGFNFYFNEYGNFRPEELGFFDGPQDNGLYSLGRFRGDK 480

Db 421 nntatfrpreteletttlGFNFYFNEYGNFRPEELGFFDGPQDNGlyslgrfrgdk 480

Qy 481 GLLPQKSTIVPAGSQFYNTFYDAALKKDIYRLNYSNTNVTGVYRFGGEYGYGSDDEFK 540

Db 481 glpqlkstivpagsqfyntfydaalkkdiyrlnysntnvtgvyrfggeygygdsddefk 540

Qy 541 RAFGENSPYKHKCNQSGIYEPVLLKYGKKRANHNSVISADFGDYFMPFASYSRTHRM 600

Db 541 rafgensptykhhcnqsgiyepvllkygkkranhnsvisadfgdyfmpfasystrhrm 600

Qy 601 PNIQEMFYSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYSRIDNY 660

Db 601 pniqemfysqigdsgvhtalkperantwqfgfntykggllkqddtlglklvgysridny 660

Qy 661 IHNVYKQWDLNGLNIPSWSVSTGLAYTIQHNRNFKDKVHKHGFLELNYDYGRFTNLISA 720

Db 661 ihnvykqwdlgnlpswsvstglaytiqhnrnfkdvkhkgfellelndygrfftnlisa 720

Qy 721 YKSTQPTNFSDESPPNNAKEDQLKQYGLSRVLSALPRDYGRLEVGTWLGKHLGLG 780

Db 721 ykstqptnfsdesppnnakedqlkqyglsvrslalprdygrlevgtwlgkhlglg 780

Qy 781 AMRYFGKIRATABERYIDGTNGNCTSNVROLGKRSTKQETTLARQPLIFDYAAYEPKK 840

Db 781 amryfgksirataveryidgtngnctsnvrolgkrstktetlarqplifdyaayepkk 840

Qy 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFPDKDEEYTCNADKTLCLNGKYGGT 900

Db 841 nlifraevknlfdriryidpldagndaatqrysfpdkdkdeevtcnadkltclngkyggt 900

Qy 901 SKSVLTNFAARGRTFLITMSYKF 922

Db 901 sksvltnfargrtflitmsykf 922

RESULT 2

AAV69380

ID AAY69380 standard; Protein; 922 AA.

XX

AC AAY69380;

XX

DT 19-JUN-2000 (first entry)

XX

DE A BASB024 outer membrane protein of N. meningitidis.

XX

KW BASB024; outer membrane protein; N. meningitidis infection;

KW bacteremia; meningitis.

XX

OS Neisseria meningitidis.

XX

PN WO200011182-A1.

PN 02-MAR-2000.

PF 13-AUG-1999; 99WO-EP05989.

XX

PR 18-AUG-1998; 98GB-0018004.

XX

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Thonnard J;

XX

DR WPI; 2000-224702/19.

DR N-PSDB; AAZ61580.

XX

PT Novel polypeptides derived from the products of the BASB024 gene of

PT Neisseria meningitidis, useful for inducing an immune response and

PT producing antibodies useful for treating meningitis -

XX

PS Claim 5; Page 83-86; 103pp; English.

XX

CC The present sequence represents a BASB024 outer membrane protein of

CC Neisseria meningitidis. The BASB024 polynucleotide sequence was

CC first identified in the Incyte Pathoseq database containing

CC unfinished genomic DNA sequence of N. meningitidis. BASB024

CC polypeptides and polynucleotides are useful for generating an

CC immune response in an animal. Antibodies specific BASB024 polypeptides

CC are useful for treating N. meningitidis infection, which causes

CC bacteremia and meningitis.

XX

SQ Sequence 922 AA;

Query Match 99.8%; Score 4894; DB 21; Length 922;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 920; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRSSFRKLPICFYLMGVMLYHHSYAEADAGRAGSEAOQVLEDVHVHAKRVPKDKKVFDTDA 60

Db 1 mrssfrlkipcfylmgvmlhyhhsyaedagragseaqlvledvhvakrvpdkkvvftda 60

Qy 61 RAVSTRQDIFKSSNLDNIVRSIPCAFTQODKSSGIYSLNIRGDSGFGFVNTWVDGIGTQT 120

Db 61 ravstrqdifkssnldnivrsipgafatqgdkssgvlslnirgdsfgfgrvntmvdgigtq 120

Qy 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVYVKGFSGSGAGINSLAGSANLRTLCGVDDVV 180

Db 121 fystsdagragssqfgasvdsnfiaagldvvyvkgfsfgsaginslagsanlrtlcvddvv 180

QY 181 QGNNTYGLLLKGLTGCTNSTKGNAMAAIGARKWLESASGVLYIGHSRSSVAQNYRVGGG 240
 Db 181 qgnntyglllkgltgtnstkgnamaaigarkwlesasgvlyghsrrtwagnyrvggg 240
 QY 241 QHIGNFGAEXLERKORYFVQEGGLKNSNGKWERDORFYWKTKYQKYNDRPOELQY 300
 Db 241 qhignfgaeylerkoryfvgqegglknsngkwerdofrywktkyqkynndrpoelqy 300
 QY 301 IEGHDKSWENLAPQYDITPIDPSLQKQASAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 360
 Db 301 ieghdkswenlapqyditpidpslqqasagnlfkleydgvfnkytaqfrdlntkigsr 360
 QY 361 KIINRNTQFNYSLSNYSANLNLTAAYNSGRQKYPKSGKFTGWLGLDKDFETYNNAKILDL 420
 Db 361 kiinrnyqfnyslsnysanlnltaaynsgrqkypksgkftgwlglldkfefeynnakildl 420
 QY 421 NNTATFRLPRETELOTTLGFNRYFNEYKGRFPEELGLFFEDGPDODNGLYSLGRFGDK 480
 Db 421 nntatfrlpreteletlglfnryfneykgrfpeelglffedgpdodnglyslgrfgdk 480
 QY 481 GLLPQKSTIVQAGSOYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTYGYGSDDEFFK 540
 Db 481 glpqkstivqagsyfnfitydaalkkdiyrlnysntvgyrfggeytygygsddefk 540
 QY 541 RAFGENSPYKKHCNQSCGIYEPVLKKGKRRANNHVSISADFGDYFMPFASYSRTHRM 600
 Db 541 rafgensptykhhcnscgiiyepvlkkygkrrannhsvsisadfgdyfmpfasysrthrm 600
 QY 601 PNIQEMVFSQIGDGVHTALKPERANTWQGFNTYKKGKLDKDDTLGLKLVGYRSRIDNY 660
 Db 601 pniqemvfsqigdgsvhtalkperantwqgfntykgkldkddtlglklvgyrsridny 660
 QY 661 IHNYGKMDLNGNIPSWSVSTGLAYTIOHRNFKDKVHKHGFELNLYDYGRRFTNLSYA 720
 Db 661 ihnygkmdlngnipswsvstglaytiohrnfdkdkvhkhgfelnlydygrfntnlsya 720
 QY 721 YQKSTQPTNFSDESAPNASKEDQLQGVGLSRVSALPRDYGRLEVTGRWLGNKLLGG 780
 Db 721 yqkstqptnfsdesapnaskedqlqgvglsrvsalprdygrlevtrwlgknllgg 780
 QY 781 AMRYFGKSIRATAERYIDGTNGGNTSNVROLGKRISIKQETTLARQPLIFDEYAAEPPK 840
 Db 781 amryfgksirataeryidgtnggntsnvrolgkriskqetltarqplifdeyaaepk 840
 QY 841 NLIFRAEKNLFDRIYIDPLDAGNDAATQRYVSSFDPKDDEEVTCTNADKTLCKNGYGGT 900
 Db 841 nlifraevknlfdrdyidpldagndaatqryvssfdpkddeevtctnadkctlcngkyggt 900
 QY 901 SKSVLTNFARGTRFLTMSYKF 922
 Db 901 sksvltnfargtrfltmsyxf 922

RESULT 3

AA69382
 ID AAY69382 standard; Protein; 921 AA.

AC AAY69382;

DT 19-JUN-2000 (first entry)

DE A BASB024 outer membrane protein of N. meningitidis.

XX BASB024; outer membrane protein; N. meningitidis infection;
 KW bacteremia; meningitis.

OS Neisseria meningitidis.

XX W0200001182-A1.

PN 02-MAR-2000.

PD 02-MAR-2000.

XX

PF 13-AUG-1999; 99WO-EP05989.
 PR 18-AUG-1998; 98GB-0018004.
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA Thonnard J;
 PI WPI; 2000-224702/19.
 DR N-PSDB; AAZ61582.
 XX Novel polypeptides derived from the products of the BASB024 gene of
 PT Neisseria meningitidis, useful for inducing an immune response and
 PT producing antibodies useful for treating meningitis -
 XX Claim 3; Page 91-95; 103pp; English.
 XX The present sequence represents a BASB024 outer membrane protein of
 CC Neisseria meningitidis. The BASB024 polynucleotide sequence was
 CC first identified in the Incyte Pathoseq database containing
 CC unfinished genomic DNA sequence of N. meningitidis. BASB024
 CC polypeptides and polynucleotides are useful for generating an
 CC immune response in an animal. Antibodies specific BASB024 polypeptides
 CC are useful for treating N. meningitidis infection, which causes
 CC bacteremia and meningitis.
 XX Sequence 921 AA;

Query Match 97.5%; Score 4779.5; DB 21; Length 921;
 Best Local Similarity 97.9%; Pred. No. 0;
 Matches 903; Conservative 7; Mismatches 11; Indels 1; Gaps 1;

QY 1 MRSSFRLKPICFYLMGVMLYHHSYAEADAGRAGSEAOIQVLEDVHVAKRVPKDKKVFDTDA 60
 Db 1 mrsfrikpicfyimgvtlyhysyaedaagragseaqiqlvedvhwkkrvpkdkkvftda 60
 QY 61 RAVSTRQDIFKSSNNLDNIVRSIFCAFTQQDKSSGIVSLNIRGDSGGFRVNTWMDGITQT 120
 Db 61 ravstrqdfkssennldnivrslpcaftqqdkssgivslnirgdsdggfrvntwmdgitqt 120
 QY 121 FYSTSTAGRAGSGSFGASVDSNFIAGLDVVKGFSGSAGINSIAGSANLRTLGVDVV 180
 Db 121 fyststadagragssqfgasvdsnfiagldvvkgfsgsaginslagsanlrtlvgddvv 180
 QY 181 QGNNTYGLLLKGLTGCTNSTKGNAMAAIGARKWLESASGVLYIGHSRSSVAQNYRVGGG 240
 Db 181 qgnntyglllkgltgtnstkgnamaaigarkwlesasgvlyghsrrswagnyrvggg 240
 QY 241 QHIGNFGAEXLERKORYFVQEGGLKNSNGKWERDORFYWKTKYQKYNDRPOELQY 300
 Db 241 qhignfgaeylerkoryfvgqegglknsngkwerdofrywktkyqkynndrpoelqy 299
 QY 301 IEGHDKSWENLAPQYDITPIDPSLQKQASAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 360
 Db 301 ieghdkswenlapqyditpidpslqqasagnlfkleydgvfnkytaqfrdlntkigsr 359
 QY 361 KIINRNTQFNYSLSNYSANLNLTAAYNSGRQKYPKSGKFTGWLGLDKDFETYNNAKILDL 420
 Db 361 kiinrnyqfnyslsnysanlnltaaynsgrqkypksgkftgwlglldkfefeynnakildl 419
 QY 421 NNTATFRLPRETELOTTLGFNRYFNEYKGRFPEELGLFFEDGPDODNGLYSLGRFGDK 480
 Db 420 nntatfrlpreteletlglfnryfneykgrfpeelglffedgpdodnglyslgrfgdk 479
 QY 481 GLLPQKSTIVQAGSOYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTYGYGSDDEFFK 540
 Db 480 glpqkstivqagsyfnfitydaalkkdiyrlnysntvgyrfggeytygygsddefk 539
 QY 541 RAFGENSPYKKHCNQSCGIYEPVLKKGKRRANNHVSISADFGDYFMPFASYSRTHRM 600
 Db 540 rafgensptykhhcnscgiiyepvlkkygkrrannhsvsisadfgdyfmpfasysrthrm 599

QY 601 PNQEMYSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
 Db 600 pnqemyfsqigdsghvhtalkperantwqgfntykkglkqddtlgklvgyrsridny 659
 QY 661 IHNYYGKWDNLGNIPSWSGTGLAYTIOHRNPKDKVHKHGFELNVDYGRFTNLSYA 720
 Db 660 ihnvygkwdngldpwsstglaytiqhrnfkdvkhghfelenydygrftnlsya 719
 QY 721 YKSTQPTNFSDESPPNASKEDQLKQGYCLSRVSLPRDYGRLEVTGRLWGLNKLTLGG 780
 Db 720 ykstqptnfsdesppnaskedqlkqgyclsrvsalprdygrlevgtwlglnkltlgg 779
 QY 781 AMRYFGKSIRATAEERYIDGTNGTGNVRLQKGRSTKQETTLARQPLIFDFYAAYPEKK 840
 Db 780 amryfgksirataeeryidgtngtgnvnrlqkgrsiktetlarqplifdfyaaypekk 839
 QY 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFDPKDDEEVTNCNADKTLGNGKYGGT 900
 Db 840 nlifraevknlfdriryidpldagndaatqryssfdpkdkdedvtcnadktlcngkygg 899
 QY 901 SKSVLTNFAGRFTLITMSYKF 922
 Db 900 sksvltnfagrftlmtsykf 921

RESULT 4
 AAY38940
 ID AAY38940 standard; Protein; 922 AA.
 XX AAY38940;
 AC AAY38940;
 DT 08-OCT-1999 (first entry)
 XX N. gonorrhoeae antigen encoded by a variant ORF133.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
 XX Neisseria gonorrhoeae.
 OS Neisseria gonorrhoeae.
 XX WO9924578-A2.
 PD 20-MAY-1999.
 XX 09-OCT-1998; 98WO-IB01665.
 PF 01-SEP-1998; 98GB-0019016.
 PR 06-NOV-1997; 97GB-0023516.
 PR 14-NOV-1997; 97GB-0024190.
 PR 18-NOV-1997; 97GB-0024386.
 PR 27-NOV-1997; 97GB-0025158.
 PR 10-DEC-1997; 97GB-0026147.
 PR 14-JAN-1998; 98GB-0000759.
 XX (CHIR-) CHIRON SPA.
 PA Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
 PI WPI; 1999-327407/27.
 DR N-PSDB; AA212354.
 DR
 XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
 PT diagnosis, treatment and prevention of infection
 PS Claim 4; Page 480; 524pp; English.
 XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
 CC and N. gonorrhoeae antigenic proteins. They are encoded by open
 CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
 CC their fragments, their nucleic acids and antibodies are used for
 CC diagnosis, prevention (as vaccines) or treatment of Neisseria
 CC infections, such as meningitis, septicaemia and gonorrhea. Both
 CC organisms are closely related. Fragments of the nucleic acids

CC are useful as hybridisation probes and antisense reagents.
 XX
 SQ Sequence 922 AA;
 Query Match 96.9%; Score 4754; DB 20; Length 922;
 Best Local Similarity 96.3%; Pred. No. 0;
 Matches 888; Conservative 16; Mismatches 18; Indels 0; Gaps 0;
 QY 1 MRSERLKPICFYLMGVMLYHHSYAEDAGRAGSEAOIQVLEDVHVAKRYPKDKKVFDTA 60
 Db 1 mrsfrlkipcfylmgvmlhhsyaedagragseaqiqvledvnhkrrypkdkkvtfda 60
 QY 61 RAYSTRQDIFKSEENLDNIVRISIPGAFQDQKSSGIYSLMIRGDSGFRYNTWVDGITQT 120
 Db 61 ravstrqdvfkseendlnivrsipgafqdqkssgiyslirgdsgrfyrntwvmdgitqt 120
 QY 121 FYSTSDAGRAGSSQPGASVDSNFIAGLDVYKGSFSGSAGINSAGSALNRLTGLVDVV 180
 Db 121 fystsdagragssqpgasvdsnfiaagldvvkgsfsgsaginsagsalnrltclgvddv 180
 QY 181 QGNNTYGLLLKGLTGTNSTKGNAMAAICARKWLESAGSVGLYGHRSRVSVAQNYRVGGG 240
 Db 181 qgnntyglldkglgtgtnstkgnamaaigarkwlesgasvglghsrsvaqrnyrvggg 240
 QY 241 QHIGNFGAELERRKQRYFVOEGGLKFNSSNGKWERDFQRPYWKTKYQKYNDDPELOKY 300
 Db 241 qhignfgeyleerrkqyfvogegllkfnagskwerdlqrqywktykkyedpqlqky 300
 QY 301 IEHDKSWRENLAPOYDITIDPSSLKQOSAGNLFKLEYDGVFNKYTAQPRDLNLTGSR 360
 Db 301 iehdkswrenlapoyditidpssllkqosagnlfkleydgvfnkytaqprdlntlgrs 360
 QY 361 KIINRNYQFNYGLSLNSYANLNTAAYNSGRQKYPKSKFTGMGLLKDFETYNNAKILDL 420
 Db 361 kiinrnyqfnyglslnoytnlntaaynsgrqkypkaktfgwllkdfetynnakild 420
 QY 421 NNTATFRLPRETELOTTLGFNFYHNEYGNRPPEELGFFDGGDDNGLYSYLGFRFGDK 480
 Db 421 nntatfrlpretelottlgnfyhneygnrpfelgffdgddngllysygrfkgdk 480
 QY 481 GLLPQKSTIVQAGSOYFNTFYDAAALKDIIYRLNYSNTVGVYRFGGEYTYGSDDEEK 540
 Db 481 glipqkstivqagsyfnctfydaalkddiylrlnystnainyrfggeytygysenefk 540
 QY 541 RAFGENSPYKKHCNOSCGIYEPVLKYYKKRANNSHVSISADFGDYFMPFASYSRTHRM 600
 Db 541 rafgenspaykehcdpscglyepvlkyykkrannhsvisadfgdyfmpfagysrthrm 600
 QY 601 PNIOEMYFSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
 Db 601 pnioemyfsqigdsghvhtalkperantwqgfntykkglkqddtlgklvgyrsridny 660
 QY 661 IHNYYGKWDNLGNIPSWSGTGLAYTIOHRNPKDKVHKHGFELNVDYGRFTNLSYA 720
 Db 661 ihnvygkwdngldpwsstglaytiqhrnfkdvkhghfelenydygrftnlsya 720
 QY 721 YKSTQPTNFSDESPPNASKEDQLKQGYCLSRVSLPRDYGRLEVTGRLWGLNKLTLGG 780
 Db 721 ykstqptnfsdesppnaskedqlkqgyclsrvsalprdygrlevgtwlglnkltlgg 780
 QY 781 AMRYFGKSIRATAEERYIDGTNGTGNVRLQKGRSTKQETTLARQPLIFDFYAAYPEKK 840
 Db 781 amryfgksirataeeryidgtngtgnvnrlqkgrsiktetlarqplifdfyaaypekk 840
 QY 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFDPKDDEEVTNCNADKTLGNGKYGGT 900
 Db 841 nlifraevknlfdriryidpldagndaatqryssfdpkdkdedvtcnadktlcngkygg 900
 QY 901 SKSVLTNFAGRFTLITMSYKF 922
 Db 901 sksvltnfagrftlmtsykf 922

PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
XX WPI: 1999-327407/27.
DR N-PSDB; AAZ12352.
XX
XX Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for
PT diagnosis, treatment and prevention of infection
PT
PS Claim 4; Page 474; 524pp; English.
XX
XX Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis*
CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open
CC reading frames (ORFs) AAZ11972-212358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
XX Sequence 888 AA;
SQ

Query Match 93.8% Score 4599.5; DB 20; Length 888;
Best Local Similarity 97.8% Pred. NO. 0;
Matches 869; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

QY 34 EAQIOVLDDVHVAKRVPKDKKVFVDARAVSTRQDIFKSSSENLNIVRSIPGAFQODKS 93
Db 1 eaqlvledvnhvakrvpkdkkvtfdaravstrqdfksssenlndnrvsipaftqgdk 60
QY 94 SGIVSLNIRGDSGFRVNTMVDGITFTYSTDAGRAGSSQFGASVDSNFTIAGLDVVK 153
Db 61 sgivslnirgdsgrvntmvdgitftystdagragsqfgasvdsnftiagldvdk 120
QY 154 GSFGSGAGINSLAGSANRLTGLVDDVVOGNTTYGILLKGLTGNTSKGNMAAIGARKWL 213
Db 121 gsfsgsaginslagsanrltglvddvvoqnttygillkgltgntskgnmaaaigarkwl 180
QY 214 ESGASGVLYGHSSRRSAQNRVYGGGQHGIGNFGAEYLERRKORYFQEGGLKFNSNGK 273
Db 181 esgasgvlyghssrrsaqnrvygggqhignfgaeylerkaryfvqegalkfnsdsk 240
QY 274 WERDFORPYWKTWYOKYNDPQELQYIEGHDKSWRENAPQYDITPIDPSSLKQOSAGN 333
Db 241 werdlgrqwkypkyknynn-qelqyieehdkswrenlpxqyditpidpsslkqqsagn 299
QY 334 LFKLEYDGVFNKYTAQFDRLNTKIGSRKTIINRYQFNGLSLNLYANLNTAAYNSGRK 393
Db 300 lfkleydgvfnkyltaqfdrlntkigsrkliinryqfnnglslnlyanlntaaynsgrk 359
QY 394 YPKGSFTGMLKIDFETYNNAKILDLNNTATPRLPRETELQTLTGTFNHFHNGKRRP 453
Db 360 ypgskftgmlkldfetynnakildnntatprlpretelqtltgtnfhnfyhngkrrfp 419
QY 454 EEGLGFDDPDQNGLYSLYGRFPGDKGLLPQKSTIVPAGSQYFNTFFDAALKKDIYR 513
Db 420 eeiglffddpdqnglyslgrfpgdkglpstkstivpagsqyfnfdaalkkdiyr 479
QY 514 LNYSTNTVGRFGEVYGYGSDDEKFRAGFENSPYKHCNQCSCGIYEPVLKYYKKRA 573
Db 480 lnystntvgrfgevgygygsddekfragfenspykhcncscgiyepvlkyykkra 539
QY 574 NNHSVTSISADFGDYFMFPFASYSYTHRMPIQEMFYFSQIGDSGVHTALKPERANTWQFGN 633
Db 540 nnhsvisadfgdyfmfpfasystrhmpniqemfysqigdsghvhtalkperantwqfgn 599

QY 634 TYKKGLLKQDDTLGLKLVGYSRIDNYIHNVYKWWDLNAGNIPSWSSSTGLAYTIQHNR 693
Db 600 tykkgllkqddtlgllklyvrsridnyihnvgykwwdngldpsvssgltaytiqhrnf 659
QY 694 KDKVHKHGFELNLYDYGRFFTNLSYAYOKSTOPTNFSASESPNNASKEDQLKQYGLS 753
Db 660 kdkvkhghfelelnlydygrfftnlsyayqkstqptnfsasespnnaskedqlkqygl 719
QY 754 RVSALPRDYGRLEVGTRWLGKNTLGGAMRYFGKSTRATAEERYIDGTNGGNTSNVRQLG 813
Db 720 rvsalprdygrlevgrtwlgnkltlggamryfgksrataeeryidgtnggntsnvrqlg 779
QY 814 KRSIKQTEIAROPLIJDFVAAVEPKKNLIFRAEVKNLFDERRVIDPLDAGNDAATQRYYS 873
Db 780 krsikqtetlarqplifdfyaayepkknlifraevknlfdrvidpldagndaatqryys 839
QY 874 SFDPKDKDEVTNCNADKTLGNGYGTGTSKSVLTNFAARGRTFLITMSYKF 922
Db 840 sfdpkdkdedvtcnadktlcngkygtsksvltnfargrtflitmsykf 888

RESULT 7
AAY38938
ID AAY38938 standard; Protein; 871 AA.
XX
AC AAY38938;
XX
DT 08-OCT-1999 (first entry)
XX
DE N. meningitidis strain A antigen encoded by a partial ORF133.
KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
KW treatment; *Neisseria* infection; meningitis; septicaemia; gonorrhea.
XX
OS *Neisseria meningitidis*.
XX
PN W0924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
XX WPI: 1999-327407/27.
DR N-PSDB; AAZ12353.
XX
XX Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for
PT diagnosis, treatment and prevention of infection
PT
PS Claim 4; Page 477; 524pp; English.
XX
XX Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis*
CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open
CC reading frames (ORFs) AAZ11972-212358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
XX Sequence 871 AA;
SQ

```
Query Match          90.0%; Score 4413; DB 20; Length 871;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 833; Conservative 3; Mismatches 35; Indels 0; Gaps 0;

QY 52 KDKKVFTRAVSTRODIFKSSNLDNIVRSIPGAFQODKSSGIVSLNTRGDSGGRVN 111
DB 1 kdkkvtadaravstrdqdfksxlenldnivrxiPGAftxqkssgsvslnirxdsgfgrvn 60

QY 112 TMVDGTTQTFYSTSTAGRAGSSQFCASVDSNFIAGLDVVKGSFSGAGINSLAGSANL 171
DB 61 tmvdgtxtfststdagragssqfgasvdsnfagldvkvksgfsgaginslagsanl 120

QY 172 RTLGVDVVGNNYTGILLKGLTGNTSGNMAAAGARKWLESASVGVLYCHSRSSVA 231
DB 121 rtlxvddvvgvngxtgylilkgltgntsgnmaaaigarxwlesasvgylyghnsrsva 180

QY 232 QNYRVGGGGHIGNFGAELERKQRYFVQEGGLKFNNSGKWERDFQRPYKTKWYQY 291
DB 181 qnyrvvggghignfgaelyerrkqryfegegglkfnnsqkwerdfqsywtkwyqy 240

QY 292 NDPQELQKYIEGHDKSWRENLAPOYDITPIDPSLSQOASAGNLFKLEYDGVFNKYTAQFR 351
DB 241 dapqelqkyieghdkswnrenlapoyditiPIDPSLSQOASAGNLFKLEYDGVFNKYTAQFR 300

QY 352 DLNWKIGSRKLIINRNQFNGLSLNSYANLNLTAAYNSGKQKYPKSKFTGWLKDFET 411
DB 301 dlntkigsrkliinrnqfnvglslnpytnlnltaaynsgrqkypkskftgwlkdfet 360

QY 412 YNNAKILDLNNTATFRLPRETELOTTLGFNFYHNEYGNRPFPEELGLFFOGPDQDGLYS 471
DB 361 ynnakildlntatfrelpretelotltlgfnfyhneygnrpfpeelglffogpdqndglys 420

QY 472 YLGRFGDKGLLPQKSTIVOPAGSQYFNTFYDAALKDILYRLNYSNTVYRFGGEYTG 531
DB 421 ylgRfgdkglpQKstivopagsyfnfTYdaalkdilyrlnysntvyrfggeyTG 480

QY 532 YGSDDEFKRAFGENSPTYKKHCNQCIGYEPVLKYGKRRANNHVSISADFGDFMPF 591
DB 481 ygSddefkRafgensptykKhcnqcigYepvlkYgkrrannhvsisadfgdfmpf 540

QY 592 ASYSRTHRMPIQEMFYFSQIGDSGVHTALKPERANTWQGFNTYKKGLLKODDTLGLKLV 651
DB 541 asysrthrmpiqemfyfsqigdsghvtalkperantwqgfntyKkglLkODDTLGLKLV 600

QY 652 GYRSRIDNYIHNHYGKWDNLGNIPSWSTGLAYTIOHRNFKDKVHKHGFELNLYDG 711
DB 601 gYrsridnyIhnhygKwdnlgnipswstglaytiOHrnfKdkvKhghfELnlyDG 660

QY 712 RFTNLSYAYOKSTQPTNFSDESNNASKEDQLKOGYGLSRVSALPRDYGRLEVGTWR 771
DB 661 rftnlsyayokstqptnfsdesesnnaskedqlkggyglsrvsalprdygrlevgtwr 720

QY 772 LSNKLTLLGGAMRYFGKSIRATAERYIDGTNGNTSNVROLGKRSIKQETFLARQPLIFD 831
DB 721 lgnkltlggamyrfgksirataeryidxtngntsnfrlgkrsixqetflarqplifd 780

QY 832 FYAAEPKMLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEEVTCTNADKT 891
DB 781 fyAAepkmlifraevknlfdriyDpldagndaatqryysfDPkDKDEEVTCTNADKT 840

QY 892 LCNKYGVTGTSKSLTNFARGRTFLITMSYKF 922
DB 841 lcnkygvtgtsksltnfargrtflitmsyKF 871

RESULT 8
AAY44428
ID AAY44428 standard; Protein: 947 AA.
XX
AC AAY44428;
XX
DT 22-MAR-2000 (first entry)
```

```
XX M. catarrhalis (ATCC 43617) BASB021 polypeptide.
DE
XX BASB021; HasR; outer membrane haem-binding protein; sinusitis;
KW otitis media; pneumonia; nosocomial infection; auditive nerve damage;
KW delayed speech learning.
XX Moraxella catarrhalis.
OS
XX WO9964602-A2.
PN
XX 16-DEC-1999.
PD
XX 31-MAY-1999; 99WO-EP03824.
PF
XX 09-JUN-1998; 98GB-0012440.
PR
XX (SMTK) SMITHKLINE BEECHAM BIOLOGICALS.
PA Thonnard J;
PI
XX MPI; 2000-116545/10.
XX N-PSDB; AA229682.
DR
XX New isolated Moraxella catarrhalis BASB021 polynucleotides, used to
PT develop products for the diagnosis, prevention and treatment of
PT infections causing e.g. otitis media -
XX Claim 1; Page 81-84; 87pp; English.
PS
XX The present sequence is BASB021 polypeptide, which shows sequence
CC homology with Serratia marcescens HasR outer membrane haem-binding
CC protein. It is encoded by DNA from Moraxella catarrhalis strain Mc2931
CC (ATCC 43617). BASB021 polynucleotides and polypeptides may be used for
CC prognosis, staging of disease, determining response to drug treatment
CC and drug screening. They can be used in vaccines for generating an immune
CC response and for preventing or treating Moraxella infections which may
CC cause otitis media, pneumonia, sinusitis, nosocomial infections and
CC invasive diseases, auditive nerve damage, delayed speech learning,
CC infection of upper respiratory tract and inflammation of the middle ear.
CC Anti-BASB021 antibodies can be used to diagnose and treat Moraxella
CC infections.
XX Sequence 947 AA;
SQ

Query Match          53.0%; Score 2598; DB 21; Length 947;
Best Local Similarity 55.4%; Pred. No. 5 3e-179;
Matches 501; Conservative 145; Mismatches 234; Indels 24; Gaps 10;

QY 39 VLEDVHVKAARR-VPKDKKVFTRAVSTRODIFKSSNLDNIVRSIPGAFQODKSSGIV 97
DB 48 ildevvvtatngtkksqkpfktkasatsrenvfnasenaialvrsvpgaftqdkksglv 107

QY 98 SLNIRGDSGGRVNTWVDGTTQTFYSTSTAGRAGSSQFCASVDSNFIAGLDVVKGSFS 157
DB 108 slnvrGdsGgransmvdtqtfyststdagrggtsqfgavidgnfiagvelnksfn 167

QY 158 GSAGINSLAGSANLRTLGVDVVGNNYTGILLKGLTGNTSGNMAAAGARKWLESASG 217
DB 168 gkgvltitfgsanfrtlnadddvdkdknfglkgkntadkhnfmlaggg9gwlnd - 226

QY 218 SVGVLYCHSRSSVAQNYRVGGGGHIGNFGAELERKQRYFVQEGGLKFNNSGKWERD 277
DB 227 sisalyayshkdisqnykvggggthignvgdillskqkvfakchaltynearsrswkd 286

QY 278 ---FQRPYKTKWYQY-----NDPQELQKYIEGHDKSWRENLAPOYDITPID 322
DB 287 ltkldketgplwdrkyfggkcygigcldtkekdfeyvadkqgqwkghakeysitpid 346

QY 323 PSSLKQOASAGNLFKLEYDGVFNKYTAQFRDLNWKIGSRKLIINRNQFNGLSLNSYANL 382
DB 347 italnqtkskshlakiirynndtsdvglrkmtdtligsrilndndnyqidaaynpneidlk 406
```


Db 567 ilhkshkkafnhsatslsaeisdyfmpfftystrhrmpniqemffsqvsnagvntalkpe 626
QY 624 RANTWQFGFNYYKGLLKQDDTLGKLGVYRSRIDNYIHNVYKQWDLNINIPSWSSYTG 683
Db 627 qsdtyqlgfntkkgflftqddvlgvklvgysrfiknyihnvvgvw--rdgmpbtwaesng 684
QY 684 LAYTIOHRNFKDKVHKHGFLELNVDYGRFETNLVSAYOKSTOPTNFSDASESNNASKE 743
Db 685 fkytiahqnykpiavksgveleinydmgrffanvysayqrtnqptnyadasprnnasqe 744
QY 744 DOLKQGYGLSRVSALPRDYGRLEVGTRWLGKNTLGCAMRYFGKSRATAEERYIDGTNG 803
Db 745 dilkqgyglsvmlpkdygrlelgrtrwfdqkltlglaaaryygsksratieeyingsr- 803
QY 804 GNTSNVRLGKRISIKOTETLARQPLIFDFVAAVEPKKNLIFRAEVKNLFDORRIDPLDAG 863
Db 804 fkntrlrenyavkktedikkpildlhvseyepikdliikaevqnlldkryvdpldag 863
QY 864 NDAATORYXSSFPDKDDEVTCNADKTLGNGKYGGTSKSVLTNFARGRTELITMSYKF 922
Db 864 ndaasqrysssl-----nnsiecaqssac-----ggsdktvlynfargrtyllsnykf 913

RESULT 10
AAY94672
ID AAY94672 standard; Protein; 918 AA.
XX AAY94672;
AC AAY94672;
XX 01-DEC-2000 (first entry)
XX Haemophilus antigen BASB070 protein sequence.
XX DE
XX KW Vaccine; BASB070; Haemophilus influenzae; strain nHi3224; pneumonia;
KW chronic bronchitis; sinusitis; otitis media; meningitis; antigen;
KW systemic disease; outer membrane protein.
XX Haemophilus influenzae.
XX OS
XX WO200050599-A1.
XX 31-AUG-2000.
XX 22-FEB-2000; 2000WO-EP01423.
XX 24-FEB-1999; 99GB-0004183.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX Ruelle J, Thonnard J;
XX WPI; 2000-543912/49.
XX N-PSDB; AAA28031.
XX Vaccinating against Haemophilus influenzae using BASB070 polypeptides
XX and/or the nucleic acids that encode them -
XX Claim 1; Fig 2; 97pp; English.
XX
XX This invention relates to a vaccine composition which contains BASB070
XX polypeptides from Haemophilus influenzae and/or the nucleotide sequences
XX that encode them. BASB070 DNA sequences encode proteins which are
XX integral outer membrane proteins with a beta-barrel conformation. The
XX invention includes BASB070 nucleotide and protein sequences from
XX H. influenzae strains RdKW20 and nHi3224. The invention includes an
XX expression vector and recombinant live microorganism comprising a BASB070
XX polynucleotide sequence, a process for the production of a BASB070
XX protein, an antibody specific for the two BASB070 proteins of the
XX invention, and a method for diagnosing an H. influenzae infection.
XX Haemophilus influenzae is a common cause of pneumonia, exacerbation of
XX chronic bronchitis, sinusitis and otitis media, and H. influenzae type b
XX causes bacterial meningitis and systemic diseases. The vaccine exhibits
XX antibacterial activity, and is used in the preparation of an agent for

CC use in generating an immunological response in a mammal.
CC The present sequence represents a BASB070 protein isolated from
CC H. influenzae strain nHi3224. The protein is used in the production of
CC the vaccine of the invention.
XX
XX Sequence 918 AA;

Query Match 50.8%; Score 2489; DB 21; Length 918;
Best Local Similarity 50.5%; Pred. No. 3.8e-171;
Matches 487; Conservative 159; Mismatches 230; Indels 88; Caps 14;

QY 1 MRSFRLKPCFCFLMGVLMVYHHYSAEDAGRAGSEAI-QVLEDDVHYKAKRVPDKKVFDT 59
Db 1 mkaikaiknlitlsinti-----gmtitgaqaeetlglidvvekvinsdkpkfte 50
QY 60 ARAVSTRQDIFKXSENLDNIVRSIPGAFTQODKSSGIVSLNIRGDSGFGFGRVNTMVDGITQ 119
Db 51 akakstrenvketqtdidqvirsipgaftqdgkgsvsvnirgenglrvntmvdqvtg 110
QY 120 TFYSTSTDRAGAGSSQFGASVDSNFIAGLDVVYKGFSGSAGINSLAGSANLRTLGLVDDV 179
Db 111 tfystaldsgsgsgsgfgaaidpnfiagvdknksnfgsgasinalagsanftlsvndv 170
QY 180 VQGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGG 239
Db 171 itddkpfgiilkgmtgsnatkmttaagrkwldngygvvygysqrevsqdyri-9g 229
QY 240 GQHIGNFGAEYLERRKORYFVQEGGLKFNSNGKWERDFORPYWKTKWQKYND----- 293
Db 230 gerlaslgdilaakekikfrndgyvl--nsagqwapdlkphwscntpslkskmsmts 287
QY 294 -----PQELQKY-----IEGHDKSWRENLAFOYDITPI 321
Db 288 ckpyrigpaatrgeilkelledgkepkdieklqksgndgteeksferrn-kdaydvap 346
QY 322 DPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNLTGSRKIINRNQYFNGLSINSVANL 381
Db 347 epsglsqsrshllkfeysddhltlgaqirfldnkisgrklienrnyqvnfnansyldl 406
QY 382 NLTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRETELQTLGFN 441
Db 407 nlmaahniqktyipkgyffagvqvadklicknvanivdinnshtflfpkeidkltlgn 466
QY 442 YFHNEYKGNRFPBELGLFFDPDQDNGLYSY--LGRFKGDKGLLPQKSTIVQPAGSOYFN 499
Db 467 yftneysknrfpeelslfyvneshdgglyslsnkgrysgskgllpqrsvilqpsgkqkfk 526
QY 500 TFYFDAALKDIIYRLNYSTNTVGRFGGEYTYGYSDDDEFKRAFGENSPYKKHCNOSCG 559
Db 527 tvyfdtalskgyihlnysvfnthyaingevygnktadk----- 565
QY 560 IYEPVLKKYKKRANNSHSVISADFGYFMPFASYSRTHRMPNIQEMYFSQIGDGVHTA 619
Db 566 ineplhkshhkafnhsatslsaeisdyfmpfftystrhrmpniqemffsqvsnagvnta 625
QY 620 LKPERANTWQFGFNYYKGLLKQDDTLGKLGVYRSRIDNYIHNVYKQWDLNINIPSWV 679
Db 626 lkpeqsdytyqlgfntkkgflftqddvlgvklvgysrfiknyihnvvgd-wsrdgvpewa 684
QY 680 SSTGLAYTIOHRNFKDKVHKHGFLELNVDYGRFETNLVSAYOKSTOPTNFSDASESNN 739
Db 685 rlngflrtlahqnyqpiavksgveleinydmgrffanvysayqrtnqptnyadasprn 744
QY 740 ASKEDOLKQGYGLSRVSALPRDYGRLEVGTRWLGKNTLGCAMRYFGKSRATAEERYID 799
Db 745 askeelkqgyglsvmlpkdygrlelgrtrwfdqkltlglaaaryygsksratieeyin 804
QY 800 GTN-GGNTSNVRLGKRISIKOTETLARQPLIFDFVAAVEPKKNLIFRAEVKNLFDRIID 858
Db 805 gsyeknttrdry--yaikkteeikkpildlhvseyepikdliikaevqnlldkryvd 862
QY 859 PLDAGNDAATORYXSSFPDKDDEVTCNADKTLGNGKYGGTSKSVLTNFARGRTELITM 918


```
SQ Sequence 753 AA;

Query Match 18.4%; Score 900; DB 22; Length 753;
Best Local Similarity 28.7%; Pred. No. 1.9e-56;
Matches 272; Conservative 130; Mismatches 299; Indels 248; Gaps 32;

Qy 16 GVML-----YHHSVAE--DAGRAGSEAIQVLEDVHVHAKRVPKDKVFTDARAVSTRQD 68
D 11 gilllscgayqsisektnskdkga-aefspl-----vsvgkttsealektgatssr-t 65
Qy 69 IFKSENLDNIVRSIPGAFTQODKSSGIVSUNIRGDSGFGRVNTMVDGIIOTFTYSTSTDA 128
D 66 tdknlqslatvrmptgtytdipqggaisvniirgmsfgfrvntmvdgtdtqsfygtstsg 125
Qy 129 GRAGGS--SQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVVQGNNTY 186
D 126 ttthgstnmmagvliidpnllvavdvtrgdssgseginalagsamrtigvddvifngnty 185
Qy 187 GLLKGLTGTNSTKGNMAAIGARK--WLESGASVGVLYGHSRRSVAQNYRVGGGQHIG 244
D 186 glrfrsvgsnglrgsmialgkgsdaftdtg-sigvmaavsgssvsnfngsg----- 239
Qy 245 NFGAEYLERRKQRYFVQBGLKFNNSGKWDRDQRPYWKTKWYQKYNDDPOELQKYIEGH 304
D 240 -----inskefydk-----ymknpksqlyk----- 261
Qy 305 DKSRENLAPOYDITPIDPSSLKOOSAGNLPKLEYGVFNKYTAQFRDLNLTIGSRKIIN 364
D 262 -----mdirpde-----fnfelsartyenkfrrdits 290
Qy 365 RNYQFNGLS--LNSYANILNTAAVNSGROKYPKGSKFTGWLKDLKDFETYNNAKILDLNNT 423
D 291 ddyvikyhtpfseidfnvtastsrngkyrdsglytfy----ktsaqnrdsaldnnt 346
Qy 424 ATRFLPRETELQTTLGFNYFHNEYKKNRFPBELGLFDDGPDQDNGLYSLGRFKDKGLL 483
D 347 srftv-adndiefmlgsklmtry-----drthsaagdpkandesi 387
Qy 484 POKSTIVQAGSQYNTFYFDAAALKDIT-----YRLNVTSTNTVGRFGGEYTYGGSDEF 539
D 388 --ennfpagsgqdisalytglktrgiweadfnlnytrni-----tg----- 429
Qy 540 KRAFGENSPTVKKHCNOS--C---GIVEPVLKYYKKRRANNHSVISIADFGDYFWPFSASY 594
D 430 -----ykpacdsrvicvpggsydidckeg-----fnpsvqlsaqvtwlpqfigy 475
Qy 595 SRTHRMPIQEMFYFSQIGDSGVHTALPERANTWQFGFNTYKGLLQDDPTLGLKLKLYGR 654
D 476 sksmrapniqemffnsaggasmnplpetaetwqagfnidtrdlveqdalrfkalayr 535
Qy 655 SRIDNYIHN-----VYKKW-----DLNGNIPSWVSSTGLAYTIQHRNF 693
D 536 sriqniysesylvcsgrkcselpvigngwegisdeysdmnyiynvsa----- 584
Qy 694 KDKVHKHGFLELANYDGRFFTNLSYAKSTQPTNDSASESPNNASKEDQLQKGYLS 753
D 585 sdvak-gfelemdydgafagrllsfqgqcdqptsasth-----fgag 628
Qy 754 RVSALPRDYGLEVGTWRGLNKLTLGGAMRYFGKSIRATAEERYIDGTNGGNTSNVRQLG 813
D 629 ditelprkymtldtgrvffdnalltltiikytgkarlspofeqdehtga----- 678
Qy 814 KRSTKQFETTLARQLIFDFYAAEPKKNLIPRAEVKNLFDPRYIDPDLADGNDAAATQRYYS 873
D 679 --iikq--dlpqiptlidiytcyeynrltllksvqnlmnrdisenaln----- 725
Qy 874 SFDPKDKDEEYTCNADKTLCKNGKYGGSKSVLTNFAGRFTPLITMSYKF 922
D 726 -mmpglgdethpans-----argrtwifggdirtf 753

RESULT 13
```

```
ABB52968
ID ABB52968 standard; Protein; 753 AA.
XX
AC ABB52968;
XX
DT 11-FEB-2002 (first entry)
XX
DE Escherichia coli polypeptide SEQ ID NO 1277.
XX
KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicemia;
KW pyelonephritis; antibiotic resistance.
XX
OS Escherichia coli.
XX
PN WO200166572-A2.
XX
PD 13-SEP-2001.
XX
12-MAR-2001; 2001WO-EP03445.
XX
10-MAR-2000; 2000FR-0003145.
PR
02-FEB-2001; 2001FR-0001449.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
WPI; 2001-550253/61.
XX
A library of DNA fragments of Escherichia coli strains for the
PT phylogenetic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -
XX
Example 6; Fig 6; 646pp; English.
XX
The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.
XX
SQ Sequence 753 AA;
```

```
Query Match 18.4%; Score 900; DB 22; Length 753;
Best Local Similarity 28.7%; Pred. No. 1.9e-56;
Matches 272; Conservative 130; Mismatches 299; Indels 248; Gaps 32;
```

```
Qy 16 GVML-----YHHSVAE--DAGRAGSEAIQVLEDVHVHAKRVPKDKVFTDARAVSTRQD 68
D 11 gilllscgayqsisektnskdkga-aefspl-----vsvgkttsealektgatssr-t 65
Qy 69 IFKSENLDNIVRSIPGAFTQODKSSGIVSUNIRGDSGFGRVNTMVDGIIOTFTYSTSTDA 128
D 66 tdknlqslatvrmptgtytdipqggaisvniirgmsfgfrvntmvdgtdtqsfygtstsg 125
Qy 129 GRAGGS--SQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVVQGNNTY 186
D 126 ttthgstnmmagvliidpnllvavdvtrgdssgseginalagsamrtigvddvifngnty 185
Qy 187 GLLKGLTGTNSTKGNMAAIGARK--WLESGASVGVLYGHSRRSVAQNYRVGGGQHIG 244
D 186 glrfrsvgsnglrgsmialgkgsdaftdtg-sigvmaavsgssvsnfngsg----- 239
```

```
Qy 245 NFGAEYLERRKQRYFVOEGGLKFNNSGKWERDFQRPYWKTKYQKYNDOELOKYIEGH 304
Db 240 -----inskefydk-----ymknpksqlyk----- 261
Qy 305 DKSRENLAPOYDITPIDPSLKOOSAGNLFKLEYDGFVNKYTAQFRDLNFKIGSRKIIN 364
Db 262 -----mdirpde-----fnfselsartyenkfrtrdits 290
Qy 365 RNYQFNGLS--LNSYANLNLTAAVNSGROKYPKGSKKTGMLLKDFETYNNAKLTLDLNN 423
Db 291 ddyiyikyhtpfseidfnvcasrqnqkyrdgslytfy-----ktsaqrnsdaldnnt 346
Qy 424 ATFRLPRETELQTTLGFNFYHNEYKGNRPPELGLFFDGDQDNGLSYILGRFKGDKGLL 483
Db 347 srftv-adndiefmglskmltry-----drtihsaagdpkangesi 387
Qy 484 POKSTIVOPAGSQVFNTFYDAALKKDI-----YRLNVTSTNVGRFGEYTYGYGSBDEF 539
Db 388 --ennpfapsqqdisalytqglkvtrglweadfnlnytrnri-----tg----- 429
Qy 540 KRAFGENSPYKKHCNOS--C--GIYEPVLKYYKKRANHSVSIADFGDYEMPASY 594
Db 430 -----ykpacdsrvicvpggsydidckeg-----fnpsvqlsaqvtwlpqfigy 475
Qy 595 SRTHRMPIQEMYSQIGDSGVHTALPERANTWQGFNTYKGLLKQDDTLGLKLVGYR 654
Db 476 sksmrapniqemffsnsgasmnflkperatwqagfnidtrdlveqdalrfkalayr 535
Qy 655 SRIDNYIHN-----VYKWW-----DLNGNIPSWSVSTGLAYTIQHRNF 693
Db 536 srignyisyesylvcsggrkcslepivngwegisdeysdmnyivnsa----- 584
Qy 694 KDKVHKGFLELNDYDGRPTNLISYAKQSTOPTNFSDASESPNNASKEDOLKQGYCLS 753
Db 585 sdvtak-gfelemdydagfagrfsfqqcdqptstasth-----fag 628
Qy 754 RVSAALPDYGRLEVTRWLGKNTLGGAMRYFGKSI RATAEERYIDGTNGGNTSNVRQLG 813
Db 629 ditelpkymtldtgvrfndnaltgtiikytgkarlspdfeqdehtga----- 678
Qy 814 KRSIKQETLARQPLIFDFYAAYPEKNLIFRAEVKNLFDPRYIDPLDAGNDAATQRYYS 873
Db 679 --likq--dlpqtliidltyeynrltlklsvqnlmordysealnkln----- 725
Qy 874 SFDPKDXDEVTNADKTLGCKYGGTSKSVLTNPFARGRTFLITMSYKF 922
Db 726 -mmpglgdethpans-----argrtwifggdfrf 753

RESULT 14
AAY07476
ID AAY07476 standard; Protein; 915 AA.
XX
AC AAY07476;
XX
DT 17-AUG-1999 (first entry)
DE N.gonorrrheae transferrin binding protein.
XX
KW Transferrin binding protein; iron; outer membrane protein; uptake;
KW Neisseria gonorrhoeae; nutrient; growth; Neisseria meningitidis; pathogen;
KW receptor; antibody.
XX
OS Neisseria gonorrhoeae.
XX
US 5912336-A.
PN
PD 15-JUN-1999.
XX
PF 23-DEC-1994; 94US-0363124.
XX
PR 23-DEC-1994; 94US-0363124.
```

```
PR 23-AUG-1990; 90US-0572187.
PR 05-NOV-1992; 92US-0973336.
PR 20-SEP-1993; 93US-0124254.
XX
XX (UYNC-) UNIV NORTH CAROLINA.
XX
XX Cornelissen CN, Sparling PF;
XX WPI; 1999-357219/30.
XX N-PSDB; AAX78929.
XX
XX Nucleic acid molecules encoding transferrin binding proteins
PT
XX Disclosure; Fig 1; 38pp; English.
XX
XX This sequence represents the transferrin binding protein from Neisseria
CC gonorrhoeae. Transferrin is an iron-regulated, outer membrane protein
CC involved in uptake of iron, an essential nutrient for the growth of
CC N.gonorrhoeae and N.meningitidis. A similar sequence from N.meningitidis
CC is shown in AAY07477. N.gonorrhoeae and N.meningitidis are two pathogens
CC of the genus Neisseria that are genetically similar, but pathologically
CC different. The growth of these cells can be inhibited by reducing the
CC ability of these cells to take up iron, e.g. by blocking the transferrin
CC receptor function. The transferrin binding proteins can be used to
CC raise antibodies for inhibiting the activity of the transferrin
XX receptor.
XX
XX Sequence 915 AA;

Query Match 5.7%; Score 278; DB 20; Length 915;
Best Local Similarity 21.4%; Pred. No. 2.2e-11;
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

Qy 5 FRLPKPCIFLMGVMLYHHSYAEDAGRAGSEAQIQVLEDVHVAK--RVPKDKKVFVDARA 62
Db 7 frlnilcslmtal---payaeav-qag-qaqekqldtqlqvakkqktrrdnevtgkl 61
Qy 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDDK--SSGIVSLNIRGDSGFRVNTMVDGIT 118
Db 62 vkadtlskeqvidrldtrydpgiavveqrgassg---ysirg-mdknrvslvtvdgla 117
Qy 119 Q-TFYSTSTDAG--RAGSSQFQASVDSNFIAGLVVVKGSFSGSAGINSIAGSANSRLTGL 175
Db 118 qiqsytaqaalgtrtagssgaineieyenvkaveiskgsnsvqsgsalagsvafqtkt 177
Qy 176 VDDVQGNNTYGLLLK-GLTGTSN--TKGNAMAA-IGARKWL--ESGASVQVLYGHSSRS 229
Db 178 addvigeqrqwigsktaysgknrgltqstalgagriggaellirtqhageirah---- 233
Qy 230 VAQNYRVGGGQGHIGNFGAEYLERRKQRYFVOEGGLKFNNSGKWERDFQRPYWKTKWYQ 289
Db 234 -----eaagrgvgsfnrlapvddgskyaayfiveeck-----nggheckanp----- 276
Qy 290 KYNDPOELQKYECHDK-----SWRENLAPOYDITPIDPSLKOOS-----AGNLF--KLEY 339
Db 277 -----kkdvvgedkrqvtstvdtytgnrfla--dplsyerswlfirpgrfrfenkrhy 326
Qy 340 DGVFNKYTAQ---PRDLNFKIGSRKII-----NRNYQFNYSLSLNSYANLN 383
Db 327 iggilerqtqtdrtdntvpafitkavfdanqkagsalrgngkyagah-----kyggl-- 379
Qy 384 TAAVNSGROKYPKGSKFTGMLLKDFETYNNAKLTLDLNNATATFRLPRETELQTTLGFNYF 443
Db 380 ---ftsgeennapvgaey-gtgvfyd-----ethtkrsrygleyv 413
Qy 444 HNEYGKNRFPPELGLFFD--GPDQDNGL-----YSYIGRFRFKGDK 480
Db 414 ytnadktdwadyarlsydrqgigldnhfqthcsadgsdkycrpsadkpfisy--yksdr 470
Qy 481 GLLPQKSTIVOPAGSQVFNTFYDAALKKDIYRLNYSNTNVGY-RFGG-----EYTCYYG 534
Db 471 viygeshkhllqaafkksfdtakl-----zhnlsvn-lgydrfgsnlrhgdy--yyq 518
```


QY 535 SDDEFKRAFGENSTYKKHCNQSGGIPEVLKYYGKKRA-----NNHSVISAD----- 583
 Db 519 san---rayslktpp-----qngkktspngreknpywvsigrgnvvtr 559
 QY 584 ----FGD-----YMPFA-----SYSRTH----- 598
 Db 560 qiclfgnntdctprsingksyyaavrdnrlgrwadvgaglrlydrsthsddgsvstg 619
 QY 599 -----RMPNIQEMYFSQIGDSGVHTALKPERANTWQ 629
 Db 620 thrtlswnagivlkpadwldltyrtstgfrlpsfaemgywrsgdikavkidpeksfne 679
 QY 630 FG-----ENTYKKGLLKQDDTLGLKLVGRSRIDN-----YIH--- 662
 Db 680 agivfkdgdfngleaswfnayrdlrv-----gyeaqikdgkeqkgnpaylnaqs 730
 QY 663 -----NVYQK--W---WDLGNIP--SWVSTGLAYT--IQHNFKDKVHKHGFLELYND 709
 Db 731 aritginilgkidwngvd---klpegwyst--faynrvrvidikkradrtidqshl-fd 784
 QY 710 YGRFTNLSTAYQKSTOPTNFSDASESPNNASKEDQLKQGYGLSRVSLPRDYGRLEVGT 769
 Db 785 -----aldpsryv-----vgsydydpegkwgvingmt-----yskakeit 819
 QY 770 RWLGNKLTGGAMRYFCKSIRATAERYIDGTNGGNTSNVRQLGKRSIKOTETLARQPLI 829
 Db 820 ellgsralling-----nsrntkatarrtrpwyi 847
 QY 830 FDEYAAYEPKKNLIFRAEVKNLFDPRYI 857
 Db 848 vdvsgyytvkhhflragvynllnhryv 875

RESULT 15
 AAY51770
 ID AAY51770 standard; Protein; 915 AA.
 XX AAY51770;
 XX
 DT 13-JUN-2000 (first entry)
 XX N. gonorrhoeae strain FA19 Tbp1 protein.
 DE
 XX
 KW Transferrin receptor; Tbp1; Tbp2; immunogenic; antibacterial; vaccine;
 KW diagnosis.
 XX
 OS Neisseria gonorrhoeae.
 XX
 PN US6015688-A.
 XX
 PD 18-JAN-2000.
 XX
 XX 07-JUN-1995; 9505-0483577.
 XX
 PR 08-NOV-1994; 9405-0337483.
 PR 08-NOV-1993; 9305-0148968.
 PR 29-DEC-1993; 9305-0175116.
 XX
 XX (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore S, Harkness R, Schryvers A, Gray-Owen S, Yang Y;
 PI Murlin A, Klein M, Chong P;
 XX
 XX WPI; 2000-181144/16.
 XX
 XX New nucleic acid encoding truncated transferrin receptor, useful for
 PT diagnosis, treatment and prevention of bacterial infections,
 PT particularly by Haemophilus -
 XX
 XX Example 6; Column 163-168; 281pp; English.
 XX
 XX This invention describes a novel isolated and purified nucleic acid (I)

CC encoding an immunogenic, C-terminally truncated analog of one of the
 CC transferrin receptor proteins Tbp1 or Tbp2 of Haemophilus influenzae
 CC which has antibacterial activity. (I) are used for recombinant
 CC production of truncated Tbp; as probes and primers for detecting, and
 CC diagnosing infection by, Haemophilus, also for isolating similar
 CC sequences from other bacteria; as immunogens for vaccinating against
 CC infections caused by bacteria that produce transferrin receptors, e.g.
 CC Haemophilus, Neisseria or Branhamella. The transferrin proteins are useful
 CC as immunogens (as above); for diagnosing infection (as antigens in
 CC immunoassays) and for raising antibodies, used for diagnosis of
 CC infections or for passive immunization. This sequence represents the
 CC transferrin receptor protein Tbp1 isolated from Neisseria gonorrhoeae
 CC strain FA19.
 XX
 XX Sequence 915 AA;

Query Match 5.7%; Score 278; DB 21; Length 915;
 Best Local Similarity 21.4%; Pred. No. 2,2e-11;
 Matches 224; Conservative 127; Mismatches 333; Indels 374; Gaps 58;

QY 5 FRLKPICFYLMGVMLYHHSYAEDAGRAGSEAOIQVLEDVHVYKAK--RVPDKKVFVTDARA 62
 Db 7 frlnlclslmtal---payaenv-qag-qakeqldtiqvakkkqktrrdnevtgikgl 61
 QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGRVNWTVDGIT 118
 Db 62 vktadtiskeqvlrdldtrdpdgiavveqgrgassg---ysirg-mdknrvsltdvdlga 117
 QY 119 Q-TFYSTSTDAG--RAGSSQFGASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLG 175
 Db 118 qiqsytaqaalgtrtagssgaineieyenvkaveiskgsnsvvegsgalagsvafqkt 177
 QY 176 VDDVVOGNNTYGLLLK-GLTGTNS--TKGNAMAA-IGARKWL--ESGASGVLYGHSRRS 229
 Db 178 addvigeqrwgiqsktaysgknrgltqslalagriggaeallirtgrhageirah---- 233
 QY 230 VAQNYRVGGGOHIGNFGAEYLERRKORYFVOEGGLKNSNGKWERDFORPYWKTWKYQ 289
 Db 234 -----eaagrgvgsfnrlapvdgskyaifiveeck-----ngghekkcaup----- 276
 QY 290 KYNDPQELQKYIEGHDK----SWRENAPQYDITPIDPSSLKQOS-----AGNLF--KLEY 339
 Db 277 -----kkdvgedkrqvtstvdtygpnrla--dplesyesrswlfrpfrfenkrhy 326
 QY 340 DGVFNKYTAQ---PRDLNWKIGSRKII-----NRNYOFNGLSLNSYANLNL 383
 Db 327 iggilerqtgtfdrdmtvpaflltkavfdanqkqagslrngnkyagnh-----kyggl-- 379
 QY 384 TAAVNSGRQYKPKSGKFTGWGLLKDFETYNNAKILDLNNTATFRLPRETELOTLGPNYF 443
 Db 380 ---ftsgennapvgaey-gtgvfyd-----ethckrsgleyv 413
 QY 444 HNEYGNRFPPEELGLFFD--GPDODNGL-----YSLGRFKGDK 480
 Db 414 ytnadkdtwadyarlsydrdgigldhfhfqthesadsdkycrpsadkpfisy---yksdr 470
 QY 481 GLLPQKSTIVQAGSOYFNTFYFDALKKDIYRLNYSTNTVGY-RFGG-----EYTGYYG 534
 Db 471 viygeshkliaqafkksfdtaki-----rhnlsvn-lgydrfgsnlrhqdyy-yq 518
 QY 535 SDDEFKRAFGENSTYKKHCNQSGGIPEVLKYYGKKRA-----NNHSVISAD----- 583
 Db 519 san---rayslktpp-----qngkktspngreknpywvsigrgnvvtr 559
 QY 584 ----FGD-----YMPFA-----SYSRTH----- 598
 Db 560 qiclfgnntdctprsingksyyaavrdnrlgrwadvgaglrlydrsthsddgsvstg 619
 QY 599 -----RMPNIQEMYFSQIGDSGVHTALKPERANTWQ 629
 Db 620 thrtlswnagivlkpadwldltyrtstgfrlpsfaemgywrsgdikavkidpeksfne 679

QY	630	FG-----FNTYKKGGLUKQDDTLGKLUVGYRSRIDN-----YIH----	662
		: :	
Db	680	agivkgdgnleaswnnayrdlvr-----gyeaqikdgkeqvgknpaynags	730
QY	663	----NVYCK-W---WDLNGNIP-SWVSSTGLAYT-IOHNFNDKVHKHFELLELYND	709
	:	: :	:
Db	731	aritginiigkidwngwd--klpegwyst--faynrvrirkkradrtidghl-fd	784
QY	710	YGRFTNLASYAQKSTQPFPNSDASESPNNASKEDQLKGQYGLSRVSAPLRDYGRLEYGT	769
	:	: : :	:
Db	785	-----aiqprryv-----vgsygdqpegkwgvngmlt-----yskakeit	819
QY	770	RWLGNKLTGGAMRYFGKSIRATABERYIDGTNGCNTSNVRLGKRSTKQTETLARQPLI	829
	:	: :	
Db	820	ellgralling-----nsrntkatarrtpwyi	847
QY	830	FDPYAAYEPKKLIFRAEVKNLFDRRI	857
	:	: :	
Db	848	vdvsgyytvtkkhftiragvylnlhnyv	875

Search completed: July 24, 2002, 08:53:15
Job time: 517 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:45:49 ; Search time 42.33 Seconds
(without alignments)
2092.945 Million cell updates/sec

Title: US-09-762-926-4
Perfect score: 4904
Sequence: 1 MRSSFLKPCIFLYMGVMLY.....SVLTNFARGRTFLITMSYKF 922

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 71: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4840	98.7	922	G81865	probable outer mem
2	4768.5	97.2	921	E81076	TonB-dependent rec
3	2514.5	51.3	913	G64110	hypothetical prote
4	1085	22.1	755	A81436	probable outer mem
5	425	8.7	891	H83218	heme acquisition p
6	435	8.7	889	C83035	hypothetical prote
7	414	8.4	830	AH0477	probable TonB depe
8	360.5	7.4	852	C98310	hasR protein (Y089
9	360.5	7.4	923	AH2972	heme receptor hasR
10	308	6.3	784	D82437	TonB receptor-rela
11	278	5.7	915	A43335	transferrin-bindin
12	269.5	5.5	908	JN0819	transferrin-bindin
13	267.5	5.5	910	C81832	transferrin-bindin
14	266	5.4	698	E82443	heme transport pro
15	262	5.4	744	B64049	outer membrane pro
16	262	5.3	766	D97634	probable outer mem
17	262	5.3	766	AF2857	outer membrane hem
18	259.5	5.3	791	F81056	hemoglobin recepto
19	257	5.2	911	JN0821	transferrin-bindin
20	256	5.2	915	F81196	transferrin-bindin
21	243.5	5.0	940	S49087	lactoferrin bindin
22	243	5.0	943	G81070	lactoferrin-bindin
23	242	4.9	912	C64107	transferrin-bindin
24	237.5	4.8	792	S61335	hemoglobin recepto
25	233	4.8	877	AC2211	heme transport pro
26	233	4.8	912	S70901	transferrin-bindin
27	228.5	4.7	911	S70911	transferrin-bindin
28	226.5	4.6	944	C81798	lactoferrin bindin
29	225.5	4.6	843	A87275	TonB-dependent rec

30	223	4.5	764	2	H83055	probable outer mem
31	222	4.5	914	2	S70906	transferrin-bindin
32	218.5	4.5	723	2	C64058	outer membrane pro
33	212	4.3	953	2	B64083	hemoglobin-binding
34	211.5	4.3	725	2	A57148	outer membrane pro
35	210.5	4.3	931	2	S66574	transferrin-bindin
36	205.5	4.2	1084	2	B64088	hemoglobin-binding
37	204.5	4.2	810	2	A81965	hemoglobin-naptog
38	203.5	4.1	720	2	S58133	Fe-regulated prote
39	195	4.0	2817	2	B97033	uncharacterized pr
40	191.5	3.9	851	2	A83484	probable heme util
41	190.5	3.9	660	2	D91176	heme utilization/t
42	190.5	3.9	660	2	E86022	outer membrane hem
43	189.5	3.9	713	2	A56268	Fe-regulated prote
44	185	3.8	714	2	F81962	probable iron-regu
45	185	3.8	795	2	C83626	probable TonB-depe

ALIGNMENTS

RESULT 1

G81865

probable outer membrane substrate binding protein NMA1700 [imported] - Neisseria meningitidis

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: G81865

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Moulton, S.; Mungall, K.; Quail, M.A.; Rajandrea; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandrea

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491

A:Reference number: A81775; MUID:20222556

A:Accession: G81865

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-922 <PAR>

A:CROSS-references: GB:AL162756; GB:AL157959; MID:g7380091; PIDN:CAB84928.1; PID:g738

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: NMA1700

Query Match	98.7%	Score	4840	DB	2	Length	922
Best Local Similarity	98.7%	Pred. No.	8.4e-299				
Matches	910	Conservative	3	Mismatches	9	Indels	0
Gaps	0						
QY	1	MRSSFLKPCIFLYMGVMLYHHSVAEDAGRAGSEAQIQVLEDVHVKAKRVPKDKKVF	60				
Db	1	MRSSFLKPCIFLYMGVMLYHHSVAEDAGRAGSEAQIQVLEDVHVKAKRVPKDKKVF	60				
QY	61	RAVSTRODIFKSSLENLDNIVRSIPGAFTQDDKSSGIYSLNIRGDSGFRVNTWVDGITQT	120				
Db	61	RAVSTRODIFKSSLENLDNIVRSIPGAFTQDDKSSGIYSLNIRGDSGFRVNTWVDGITQT	120				
QY	121	FYSTSTDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV	180				
Db	121	FYSTSTDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV	180				
QY	181	QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRVSQAQNVGVGGG	240				
Db	181	QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRVSQAQNVGVGGG	240				
QY	241	QHIGNFGAEYLERRKQRYFVQEGGLKFNSSGKWERDFORPYWKTKYQKYNPQELQKY	300				
Db	241	QHIGNFGAEYLERRKQRYFVQEGGLKFNSSGKWERDFORPYWKTKYQKYNPQELQKY	300				
QY	301	IEGHDKSWRENLAPOQDITPIDFSSLLKQSQAGNLFKLEYDGVFNKYTAQFDLNTKIGSR	360				
Db	301	IEGHDKSWRENLAPOQDITPIDFSSLLKQSQAGNLFKLEYDGVFNKYTAQFDLNTKIGSR	360				
QY	361	KIINRNQFNGVSLNSYANLNLTAAYNSGRQKYPKGSKFTGWLKDDFTYNNAKLTDL	420				
Db	361	KIINRNQFNGVSLNSYANLNLTAAYNSGRQKYPKGSKFTGWLKDDFTYNNAKLTDL	420				

```
QY 421 NNTATRLPRETELQTTLGFNHNEYGNKRFPEELGLFFDGGDDQDNGLSYLGREFKGDK 480
Db 421 NNTSTRLPRETELQTTLGFNHNEYGNKRFPEELGLFFDGGDDQDNGLSYLGREFKGDK 480
QY 481 GLLPQKSTIVQAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 540
Db 481 GLLPQKSTIVQAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 540
QY 541 RAFGENSPYKHKCNQSCGIEYEPVKKYKGRANNSHVSISADFGDYFMPFASYSRTHRM 600
Db 541 RAFGENSPYKHKCNQSCGIEYEPVKKYKGRANNSHVSISADFGDYFMPFASYSRTHRM 600
QY 601 PNQEMTFYSGISGVHTALKPERANTWQFGFNTYKGLLKQDDTGLKLVGYSRSDNY 660
Db 601 PNQEMTFYSGISGVHTALKPERANTWQFGFNTYKGLLKQDDTGLKLVGYSRSDNY 660
QY 661 IHNVYKQWDLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNNDYGRFETNLSYA 720
Db 661 IHNVYKQWDLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNNDYGRFETNLSYA 720
QY 721 YOKSTQPTNFSADSESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVTGRWLNKLTGG 780
Db 721 YOKSTQPTNFSADSESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVTGRWLNKLTGG 780
QY 781 AMRYFGKSIATAEERYIDGTNGTNSNROLGKRSIKQETLARQPLIFDFFAAAYEPKK 840
Db 781 AMRYFGKSIATAEERYIDGTNGTNSNROLGKRSIKQETLARQPLIFDFFAAAYEPKK 840
QY 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDEEVTCKNADKTLGNGKYGCT 900
Db 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDEEVTCKNADKTLGNGKYGCT 900
QY 901 SKSVLTNFAARGRTFLITMSYKF 922
Db 901 SKSVLTNFAARGRTFLITMSYKF 922

RESULT 2
E81076
TonB-dependent receptor NMB1497 [imported] - Neisseria meningitidis (strain MC58 serogroup B)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: E81076
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000.
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venturi, A.; Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: E81076
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-921 <TET>
A:Cross-references: GB:AE002499; GB:AE002098; MID:g7226737; PIDN:AAF41853.1; PID:g7226737
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1497
```

```
Query Match 97.2%; Score 4768.5; DB 2; Length 921;
Best Local Similarity 97.7%; Pred. No. 2.9e-294;
Matches 901; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

QY 1 MRSSFLKPCIFYLMGYMLYHSHVAEDAGRAGSEAOIQVLEDVHVKAKRYPKDKKVFDTA 60
Db 1 MRSSFLKPCIFYLMGYMLYHSHVAEDAGRAGSEAOIQVLEDVHVKAKRYPKDKKVFDTA 60
QY 61 RAVSTRQDIFKSSNLDNIVRSIPGAFQDQKSSGIVSLNIRGDSGFRVNTMWDGTTQT 120
Db 61 RAVSTRQDIFKSSNLDNIVRSIPGAFQDQKSSGIVSLNIRGDSGFRVNTMWDGTTQT 120
```

```
QY 121 FYSTSDAGRAGSSQFGASVDSNFTAGLDVVKGSFSGSAGINSLAGSANLRTGLVDDVV 180
Db 121 FYSTSDAGRAGSSQFGASVDSNFTAGLDVVKGSFSGSAGINSLAGSANLRTGLVDDVV 180
QY 181 QGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESASGVLYGHSHRRSVAQNVYRGVGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESASGVLYGHSHRRSVAQNVYRGVGGG 240
QY 241 OHIGNEGAELERRKORYFYQEGGLFNSNGKWERDFORPYWKTWKYQKYNPQELQKY 300
Db 241 OHIGNEGAELERRKORYFYQEGGLFNSNGKWERDFORPYWKTWKYQKYNPQELQKY 300
QY 301 IEHSDKSWRENLAPOYDITPIDPSSLKQOAGNLKLEYDGVFNKNTAQPRDLNTKIGSR 360
Db 301 IEHSDKSWRENLAPOYDITPIDPSSLKQOAGNLKLEYDGVFNKNTAQPRDLNTKIGSR 360
QY 361 KIINRNYQFNYSGLNSYANLNLTAAVNSGRQYKPKSGKFTGMLLKDFETYNNAKILDL 420
Db 361 KIINRNYQFNYSGLNSYANLNLTAAVNSGRQYKPKSGKFTGMLLKDFETYNNAKILDL 420
QY 421 NNTATRLPRETELQTTLGFNHNEYGNKRFPEELGLFFDGGDDQDNGLSYLGREFKGDK 480
Db 421 NNTATRLPRETELQTTLGFNHNEYGNKRFPEELGLFFDGGDDQDNGLSYLGREFKGDK 480
QY 481 GLLPQKSTIVQAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 540
Db 481 GLLPQKSTIVQAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 540
QY 541 RAFGENSPYKHKCNQSCGIEYEPVKKYKGRANNSHVSISADFGDYFMPFASYSRTHRM 600
Db 541 RAFGENSPYKHKCNQSCGIEYEPVKKYKGRANNSHVSISADFGDYFMPFASYSRTHRM 600
QY 601 PNQEMTFYSGISGVHTALKPERANTWQFGFNTYKGLLKQDDTGLKLVGYSRSDNY 660
Db 601 PNQEMTFYSGISGVHTALKPERANTWQFGFNTYKGLLKQDDTGLKLVGYSRSDNY 660
QY 661 IHNVYKQWDLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNNDYGRFETNLSYA 720
Db 661 IHNVYKQWDLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNNDYGRFETNLSYA 720
QY 721 YOKSTQPTNFSADSESPNNASKEDQLKQ 780
Db 721 YOKSTQPTNFSADSESPNNASKEDQLKQ 780
QY 781 AMRYFGKSIATAEERYIDGTNGTNGT 840
Db 781 AMRYFGKSIATAEERYIDGTNGTNGT 840
QY 841 NLIFRAEVKNLFDRIYIDPLDAG 900
Db 841 NLIFRAEVKNLFDRIYIDPLDAG 900
QY 901 SKSVLTNFAARGRTFLITMSY 922
Db 901 SKSVLTNFAARGRTFLITMSY 922

RESULT 3
G64110
hypoetical protein H11217 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: G64110
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, R.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995.
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: G64110
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
```


Qy 725 TQPTNFSDASESPNNASKEDQKQGYGLSRYSALPRDYGRLEVGTRWLGKRLTLGGAMRY 784
Db 594 M-----YSRQDTSSTISQTSGLLGSFSASKIMELPKDYANVELGFR-LNDKISFGGIAY 648
Qy 785 FGKSRATAEERYIDGNGGHTS-NVRQLGKRSTKQETTLARQLIIFDFAAYEPKKNLI 843
Db 649 TGKA-----KRVAPNTDDWKNKDPNNPYPKPT---TQDLPKPIIIVDLWNIEWFNLT 699
Qy 844 FRAEVKNLFDRRYIDPLDAGNDAATQRYSSFDPKDKDEEVTGNADKTLCLNGKYGGTSKS 903
Db 700 MRAEVQNLFDKNYMDALNAYNSLONQLOY-----NG--AGDPIY 736
Qy 904 VLTNFARGRTFLITMSYKF 922
Db 737 LFSNSARGRTFIVSFYKY 755
RESULT 5
heme acquisition protein HasR PA3408 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83218
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Braham, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, S.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337
A:Accession: H83218
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-891 <STO>
A:Cross-references: GB:AE004762; GB:AE004091; NID:g9949544; PIDN:AAG06796.1; GSPDB:GN001
C:Genetics:
A:Experimental source: strain PA01
A:Gene: hasR; PA3408

Query Match 8.7% Score 425; DB 2; Length 891;
Best Local Similarity 22.2%; Pred. No. 6.5e-19;
Matches 215; Conservative 114; Mismatches 382; Indels 238; Gaps 37;

Qy 30 RAGSEAOIQVLEDVHVKAKRPDKKVFETDARAVST--RODIFKS-SENLDNIVRSIPGA 86
Db 125 RVSQDILLVQMSPSV-ISAAR--PDDVVVQTPHPSVSVIGREQIERNPPRHAADMLEETPGV 181
Qy 87 FTQODKSSGIVSLNIRGDSGFRVNTWVDGITQTFYSTDAGRAGSSQFGA-SVDSNF 145
Db 182 YSSVSQDDPGLSVNIRGTQDYGRYNMSVDGMRQNYQ-----QSGHQQRNGTLVYDPEL 234
Qy 146 IAGLDVYKGSFSGSAGINSLAGSANLRTGLGVDDVVQGNNTYG---LLKGLTG-TNSTK- 200
Db 235 LSEVVIDKGASSAMGAGVIGGIANTFLRDLVRCQGVGRVLTSLGGDANGTHF 294
Qy 201 -GNMAAIGARKWLESGASGVLYGHSSRVSQAQRYRVGGGQHTGFGAEYLERRKORYF 259
Db 295 IGSAFAAIGTEVW-----DMLVAASERHLG-DYDPGTKGS-IG-----ELRTCAWF 338
Qy 260 VQEGGLKFNNSGKWERDQFQPYWKTWYQKYNPDQELQKYIEGHKSWRENLAPOVDIT 319
Db 339 NPEAGQVRK-----HSPVAYSGVY-----MRSRLAKLGVAL 369
Qy 320 PIDFSSLLKQOSAGNLFKLEVDGVFNKYTAQFRDLNTKIGSRKLIINRYQFNGLSL-NSY 378
Db 370 PQDQ---RLQFSYLTQVSSYDDA-NMLNTENQALWEKLGSSDVRQAQFAIDYGVAPNPL 425
Qy 379 ANLNLTAAY--NSGRQKYPKSGFTGGLLKDFETYNNAKILDLNNTATFELPRETELQ 436
Db 426 VDFRAKLYYVDNRNRQOTLQRTGTPGYSIYQTDYQ-----AQAOQNTSTALDLSFLRA 481
Qy 437 TLGNFYHNEYGKNR-----PPEELGLFFDGDQDQNGLYSYLGRFKGDKGLLPQ 485

Db 482 NYGLEFFYDKVRPDSQPRASTSAVGFPAAEGM---TPKGDRLALGSLFARLDYD----- 532
Qy 486 KSTIVQAGSOYNTFFYDAALAKDIYRL-----NYSTNTVGYRFGGEYTYGYSDESDEF 539
Db 533 -----YDDWLNINAGLRDYRLRGDTGFGNARTILGTTRQTDMPQLQAVD--- 578
Qy 540 KRAFGENSPTYKKHCNCSOGIYEPVKYKGRANNHSHVSISADFG-DYFMPFASYSRTH 598
Db 579 -REEGRFSPTF-----GLSVKPGVDWLQLFATYCKGW 609
Qy 599 RMENIQMYESQIDSG-----VHTALKPERANTWQFGNTYKGLLKQDDTLGLKLVG 652
Db 610 RPPAVTESLITGRPHGGGAENMYFNPFLSPERSKAWGVFNVLKENLWFSDRLGLKVA 669
Qy 653 YRSRIDNYIHNVYKQWDLNGNIPSWSSVSTGLAVTIQHRNFKDKVHKHGFELNLYDGR 712
Db 670 FDRVDDFIEMGMQ-----PPGYGMAGIGNSAVVNNL-DSTRFRGEYQIDYDAGL 721
Qy 713 FFTNLSYAYQKS-----TQPTNFSDASESPNNASKEDQKQGYGLSRV--- 755
Db 722 AYGQLSYTHMIGSNDFCSTAWLGGVQTQVKGSGRRPPVIDMRPDEQANAATHCSAVLGS 781
Qy 756 -SALPRDYGRLEVGTRWLGKRLTLGGAMRYFGKSIRATAERYIDGTNGGNTSNVRQLGK 814
Db 782 AEHPMDRG-----SLTLG--MRFDRDLVDGARARYSEGYSVAGGATVSQAGV 828
Qy 815 RSIKQETTLARQLIIFDFAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSS 874
Db 829 YPADWKEY-----TVYDLYGSYRVSDDELTLRLAMENTDRLAYVLPG----- 870
Qy 875 FDPKDKDEEVTGNADKTLCLNGKYGGTSKSVLT-NFARGRTFLITMSYKF 922
Db 871 -----DVLFTLGRGRTLQGTLEYQF 891
RESULT 6
C83035
hypothetical protein PA4897 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83035
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337
A:Accession: C83035
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-989 <STO>
A:Cross-references: GB:AE004902; GB:AE004091; NID:g9951162; PIDN:AAG08282.1; GSPDB:GN001
C:Genetics:
A:Experimental source: strain PA01
A:Gene: PA4897

Query Match 8.7% Score 425; DB 2; Length 989;
Best Local Similarity 23.2%; Pred. No. 7.5e-19;
Matches 230; Conservative 130; Mismatches 382; Indels 250; Gaps 42;

Qy 30 RAGSEAOIQVLEDVHVKAKRPDKKVFETDARAVSTQDIFKSSSENLD-----NIVR 81
Db 149 RMSGEAPADLSPPVVSAAELADPKETTYT-----APSSVYLSSEDDIDRFRVSVGDLQ 203
Qy 82 SIGCAFTQODKSSGIVSLNIRGDSGFRVNTWVDGITQTFYSTDAGRAGSSQFGASV 141
Db 204 GIFGVQVDSRNGGALDINIRGQGSRAVRVVDGAEQ-----ALDVIYRGYAGTQQRASY 258
Qy 142 DSNFIAGLDVVKGSFSGSAGINSLAGSANLRTGLGVDDVVQGNNTYGLLLKGLTCTNSTKG 201
Db 259 DPLVSSVTVDKRGPSTRSGAI---GGSEVMRTIGVKDILVDGKDLGVFRFTGWNN--- 312


```
Qy 908 FA--RGRTFLTMSYKF 922
Db 814 FAPSRGRTIOGGFEYKF 830

RESULT 8
C98310
hasR protein (Y08983) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: C98310
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurolo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: C98310
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-852 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK90005.1; PID:g15159974; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2875
A:Map position: linear chromosome

Query Match 7.4%; Score 360.5; DB 2; Length 852;
Best Local Similarity 21.1%; Pred. No. 7.5e-15;
Matches 201; Conservative 117; Mismatches 332; Indels 303; Gaps 37;

Qy 62 AVSTRODIFKSS-ENLDNIVRSIPGAFTQODKSS-GIVSLNIRGDSFGFRVNTWVDGITQ 119
Db 111 SVWSREAIQSAGVNRTRDFNRVSGVYAGEGNGSFPTVSPNVRGLQESGRVVSIDGARQ 170

Qy 120 TF---YSTSTADGACGSSQFGAS-----VDSNFIAGLDVYVKGSGFSAGINSLAGSANL 171
Db 171 NAQRCYST-----GGASIYSANNGQAYVDAAFIRAVEVEKMTSATSGNAGSLGGKVEF 223

Qy 172 RTLGVDVVOGNNYTYGLLLKGLTGTN--STKGNAAGAAIGARKWLESAGSVGLYCHSRRS 229
Db 224 RTVSAADLIPEGANKGAENVNRSNGYDFOGSVLAAV-----REPDPGLSEFVAGYS-RT 277

Qy 230 VAQNRVGGGGQHI-----GNFG---AEYLERKQRVYFVQ--- 261
Db 278 IMDEYKIGTKGEALSTALTMTKDLLGRDGSWTFPFKSEGDGFDGVQVTSLSWMHQNDVFOGAS 337

Qy 262 ---EGGLKFNNSGK--WERDFORPYWTKYOKYNDPQELQYIEGHDKSWRENLAPO 315
Db 338 TVVDREGVRNDSIVAKLDWDPESELIDFKSSLWLDNNTHLRTHEL-----AART 384

Qy 316 YDI-TPID-----PSSLKOOSAGNLFKLEYDGVFNKYTAQ--FRDLNTKIGSRKIINRY 367
Db 385 YAVETNLDMLGRSFGSLENTSR--FDTRAGALSINYGAEAFRDIATSVATSIQNP 441

Qy 368 QFNGLSLNSYANL-----NLTAAYNSGRKYPKSGKFTG-----WGLLKDFEYNNAK 416
Db 442 SF-----ASSYTSFSPARRRDVASLFLNGELEPADWITLSGGVRYDWSRLKGSATYTSFK 496

Qy 417 ILDLNNTATFRLPRETELQTTLGFNFHNEYKGNRPPEELGLFPDQDQNGLYSYLGRF 476
Db 497 -----ESIVTTSVPCDLVRNH----- 513

Qy 477 KDGKLLPQKSTIVOPAGSOYNTFYDAAALKDIDYRLNYSNTVGYRFGGEYTYGSGD 536
Db 514 -----TALEYFNQVFLPALPNVNASRYNVLASIW----- 543

Qy 537 DEFKRAFGENSPYKHKHCNCSGIEPVILKYKGRANNHSVSIADFG----- 585
Db 544 -----PRTSANCMPGTGI-----TTKTPVTEYPSHEVDIDRTYSAWLPSATIEFK 588

Qy 586 --DYFMFPASYSRTRHPNIOEMYF--SQIGDS-----GVHTALKPERANTWQGFNTYK 636
Db 589 PVDWFRPVVSYQSLSRPTILETFEAGARPDGDSAGYAPNQSLRAEKATYIEIGANNSF 648
```

```
Qy 637 KGLLKQDDTLGLKLVGYSRIDNYI-----HNVYKGWMDLNGNIPSWSVSTGLATIQ 689
Db 649 DGVLLDDDTLRKMAAFREVKVDYIALGYLVTDQVFDITY-----TSFVNLDDTTY--- 699

Qy 690 HRNFKDKVHKHGFPELELYDYGRFETNLISAYIQKSTOPTNFSDESASPNNAKSKEDQLKQG 749
Db 700 -----MRGLENGYDARSEWIGSGSATVLKTEWPEKTQVFSNSTTTSGE----- 744

Qy 750 YGLSRVSALPRDYGRLEVGTRWLNKLTGLGAMRYFGKSIIRATAEERYIDGTNGNTSNV 809
Db 745 -----IVANPGD-----VAPRM---KITLDGGMRFDEKFSGLARLNHVPTQSRITLDT 791

Qy 810 RQLKRSIKQITETIARQPLIFDFYAAVEPKKNLIFRAEVKNLFRDRYIDPLDAGNDAATQ 869
Db 792 GNL--REITDPYTTV-----DLVGSYAFNDKATLRAVNNLTDRKYIPASA----- 836

Qy 870 RYSSFPDKDKDEEVTCTNADKTLGNGYGTGTSKSVLNFARGRFTLTMSYKF 922
Db 837 --YT-----APGRTFIATMNVKF 852
```

RESULT 9

AH2972

heme receptor hasR [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AH2972
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; MCCI
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH2972

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-923 <KUR>

A:Cross-references: GB:AE008689; PIDN:AAL44198.1; PID:g17741777; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: hasR

A:Map position: linear chromosome

Query Match 7.4%; Score 360.5; DB 2; Length 923;
Best Local Similarity 21.1%; Pred. No. 8.4e-15;
Matches 201; Conservative 117; Mismatches 332; Indels 303; Gaps 37;

```
Qy 62 AVSTRODIFKSS-ENLDNIVRSIPGAFTQODKSS-GIVSLNIRGDSFGFRVNTWVDGITQ 119
Db 182 SVWSREAIQSAGVNRTRDFNRVSGVYAGEGNGSFPTVSPNVRGLQESGRVVSIDGARQ 241

Qy 120 TF---YSTSTADGACGSSQFGAS-----VDSNFIAGLDVYVKGSGFSAGINSLAGSANL 171
Db 242 NAQRCYST-----GGASIYSANNGQAYVDAAFIRAVEVEKMTSATSGNAGSLGGKVEF 294

Qy 172 RTLGVDVVOGNNYTYGLLLKGLTGTN--STKGNAAGAAIGARKWLESAGSVGLYCHSRRS 229
Db 295 RTVSAADLIPEGANKGAENVNRSNGYDFOGSVLAAV-----REPDPGLSEFVAGYS-RT 348

Qy 230 VAQNRVGGGGQHI-----GNFG---AEYLERKQRVYFVQ--- 261
Db 349 IMDEYKIGTKGEALSTALTMTKDLLGRDGSWTFPFKSEGDGFDGVQVTSLSWMHQNDVFOGAS 408

Qy 262 ---EGGLKFNNSGK--WERDFORPYWTKYOKYNDPQELQYIEGHDKSWRENLAPO 315
Db 409 TVVDREGVRNDSIVAKLDWDPESELIDFKSSLWLDNNTHLRTHEL-----AART 455

Qy 316 YDI-TPID-----PSSLKOOSAGNLFKLEYDGVFNKYTAQ--FRDLNTKIGSRKIINRY 367
```


A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-915 <COR>

A;Cross-references: GB:M96731; NID:g150360; PIDN:AAA25503.1; PID:g150361

A;Note: sequence extracted from NCBI backbone (NCBIN:112950, NCBIPI:112951)

C;Superfamily: bacterial pathogen transferrin-binding protein; tonB-dependent receptor

C;Keywords: membrane protein

F;75-217/Domain: tonB-dependent receptor amino-terminal homology <TNN>

F;585-915/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 5.7%; Score 278; DB 2; Length 915;

Best Local Similarity 21.4%; Pred. No. 1.4e-09;

Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

```

Qy 5 FRLPKPICFYLMGVMLYHHSYAEDRAGSEAOIQVLEDDVHVAK--RVPKDKKVFETDARA 62
Db 7 FRMLNLCLSLMTAL---PAYAENV-QAG-QAQQKQDQDTIQVAKKQKTRDRNEVTGLGKL 61
Qy 63 VSTRQDIFKSSSE-NLDNIVRSIPG-AFTQDQK--SSGIVSLNIRGDSGFRGVRNTMVDGIT 118
Db 62 VKTADTILSKQVLDIRLTRYDPCGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117
Qy 119 Q-TFYSTSTADG--RAGSSQFQASVDNFTAGLDVVKVGFSGAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEIYENVKAVEISKGSNSVQSGGALAGSVAFOTKT 177
Qy 176 VDDVVOGNTYGLLLK-GLTCTNS--TKGNAMAA-IGARKWL--ESGASGVLYGHSRRS 229
Db 178 ADDVIGEGROWIQSKTAYSGKNRGLTQSTALAGRIGAGALLIRTRHAGEIRAH---- 233
Qy 230 VAQNYRVGGGQHHIGNFGAEYLERRKQRYFVQEGGLFNSNGKWERDFQRPYWKTKWYQ 289
Db 234 -----EAGRGVQSFNRLAPVDDGSKYAFIVEBECK----NGGHECKAMP----- 276
Qy 290 KYNDPQELQKYLEGHDK-----SWENLAPQYDITPIDSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KDDVGEDRKQVTVSTRDYTGPNRFLA--DPLSYESKSWLFRPGFREENRHHY 326
Qy 340 DGVFNKYTAQ---FRLNLTIGSKRII-----NRNYQFNGLSLNSYANLNL 383
Db 327 IGGILERTQQTFTDRMTVPFAFLTKAVFDANQKQAGSLRNGKYAGNH-----KYGGL-- 379
Qy 384 TAAVNSGRQYKPKSGKTGGLGLKDFETYNNAKLIDLNNATFRLPRETELQTLTGNFYF 443
Db 380 ---FTSGENNAVPAGAEY-GTGVFYD-----ETHTKSYRGLEYV 413
Qy 444 HNEYGNRPPEELGLFPD--GPQODNGL-----YSYLGREFKGDK 480
Db 414 YTNADKDTWADYARLSVDROGIGLDNHFQOQTHCSADGSKYCRPSADKPFY----YKSDR 470
Qy 481 GLLPQKSTIVQAGSQYFNTFYDAALKKDIYRLNYSNTVGY-RFGG-----EYTGYYG 534
Db 471 VIYGESHKLLQAAFKKSFDTAKI-----RHNLSVN-LGYDRFGSNLRHQDY--YYQ 518
Qy 535 SDDEFKRAFGENSTYKHKHCQSCGIIYEPVLKYYGKKRA-----NNHVSISAD----- 583
Db 519 SAN---RAYSLKTPP-----QNNGKKTSPNGREKNKPYWVWSIGRGNVTR 559
Qy 584 ---FGD-----YEMPEFA-----SYSRTH----- 598
Db 560 QICLFGNTYDCTPRSINGKSYAAVRDNRVLRGWADVAGLRYDRYRSTHSDGCVSTG 619
Qy 599 -----RMPNIQCYMFQIGDSGVHTALKPERANTWQ 629
Db 620 THRTLNNAGIVLKPADWLDLTYSRTGFRPLPSFAEMYGWRSGDKIKAVKIDPEKSENKE 679
Qy 630 FG-----FNTYKGLKQDDTLGLKLVGYRSRIDN-----YIH--- 662
Db 680 AGIVFKGDFGNLEASWFNNAYRDLIVR-----GYEAQIKDGKQVKGKGNPAPYLAQOS 730
Qy 663 -----NVYCK--W---WDLNGNTP-SWVSTGLAYT-IQHRNPKDKVHKHGFELNLYD 709

```

```

Db 731 ARITGINILGKIDWNGVMD---KLPEGWYST--FAYNRVRVRDIKKRADRTDIQSHL-FD 784
Qy 710 YGRFTNLNAYQKSTOPTNFSDESPPNNAKEDQLKQGYGLSRVSLPRDYCRLEVGT 769
Db 785 -----AIQPSRYV-----VGSYGQDQPEGRKGVNGMLT-----YSKAKEIT 819
Qy 770 RWLGNKLTLLGAMRYFGKSRATAEERYIDGTNGNGTNSVRLQLGKRSIKQETELARPLI 829
Db 820 ELLGSRALNG-----NSRNTKATARTRPWYI 847
Qy 830 FDFYAAEYEPKKNLIFRAEVKNLFDPRRYI 857
Db 848 VDSGYTYTVKKHFTLRAGVYNLLNHRVY 875

```

RESULT 12

JN0819

transferrin-binding protein 1 precursor - Neisseria meningitidis (strain B16B6)

C;Species: Neisseria meningitidis

C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 26-Aug-1999

C;Accession: JN0819; PN0634; S33154

R;Legrain, M.; Mazarin, V.; Irwin, S.W.; Bouchon, B.; Quentin-Millet, M.J.; Jacobs, E

Gene 130, 73-80, 1993

A;Title: Cloning and characterization of Neisseria meningitidis genes encoding the tr

A;Reference number: JN0818; MUID:93345825

A;Accession: JN0819

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-908 <LEG1>

A;Cross-references: EMBL:Z15129; NID:g297042; PIDN:CAA78831.1; PID:g297044

A;Accession: PN0634

A;Molecule type: protein

A;Residues: 25-42 <LEG2>

C;Genetics:

A;Gene: tbp1

C;Superfamily: bacterial pathogen transferrin-binding protein; tonB-dependent receptor

C;Keywords: iron transport; membrane protein; metal binding; receptor

F;1-24/Domain: signal sequence #status predicted <SIG>

F;25-908/Product: transferrin-binding protein 1 #status predicted <MAT>

F;75-217/Domain: tonB-dependent receptor amino-terminal homology <TNN>

F;578-908/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 5.5%; Score 269.5; DB 2; Length 908;

Best Local Similarity 21.0%; Pred. No. 4.8e-09;

Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;

```

Qy 5 FRLPKPICFYLMGVMLYHHSYAEDRAGSEAOIQVLEDDVHVAK--RVPKDKKVFETDARA 62
Db 7 FRMLNLCLSLMTAL---PAYAENV-QAQQKQDQDTIQVAKKQKTRDRNEVTGLGKL 61
Qy 63 VSTRQDIFKSSSE-NLDNIVRSIPG-AFTQDQK--SSGIVSLNIRGDSGFRGVRNTMVDGIT 118
Db 62 VKSSDTLSKEQVLNIRLTRYDPCGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGVS 117
Qy 119 Q-TFYSTSTADG--RAGSSQFQASVDNFTAGLDVVKVGFSGAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEIYENVKAVEISKGSNSSEYNGALGSAVAFQTK 177
Qy 176 VDDVVOGNTYGLLLK-GLTCTNSTKGNAMAAIGARKWLESAGSVGLYGHRSRVAQNY 234
Db 178 AADIIIGSGKQWQTSKTAISGKHQHALTQSLALAGR---SGGAELLIYTKRRGREHTAH 233
Qy 235 RVGGGGQHIGNFGAEYLERRKQ-----RYFVQE---GGLKFNSEN-----SGKWERDF 278
Db 234 KDAGKG--VQSFNRLVDEDKKEGGSQYRYFIVEECHNGHYAACNKLKEDASVKDER-- 289
Qy 279 QRPYWKTKWYQKYNDOQE-LQYIEGHDKSWRENLAFOYDITPIDSSLKQOSAGNLFKL 337
Db 290 -----KTVSTQDYTGSNRLLANPLEYGSQSWL--FRPGW-----HLDNR 326
Qy 338 EYDGVFNKYTAQFRLTLTKIGSRKLIINRNYQFNGLSNYANLNTAAVNSGRQKYPK- 396

```

Db	456	YAVETNLDMLRSPGSGLENTSR---	FDTRAGALSLNYGAEAPRDIAISVATSATIAQNP	512		
Qy	368	QFNVGLSLNSYANI-----	NLTAAVNSGRQYKPGSKFTG-----	WGLLKDFETYNNAK	416	
Db	513	SF-----	ASSYTSFSPAGRRDVASLFLNGELEPADMTLTSGGVRYDMSRLKGSATYYSEK	567		
Qy	417	ILDNNATATRLPRETELOTLTGFNYPHNEYGKNRPFEEELGLFDGPDQDNGLSYLGRF	476			
Db	568	-----	ESTVTTSPCDLVRNH-----	584		
Qy	477	KGDKGLELPQKSTIVOPAGSOVFNTEFYDAALKKDIYRLNSTNTVGRFEGGEYTYGSD	536			
Db	585	-----	TALSYFNQVFLPANLPWASRYNVFLASIM-----	614		
Qy	537	DEFKRAFGSPYKHKNCOSCGIYEPVLKYGKKRANHSVTSADFG-----	585			
Db	615	-----	PRTSANCMPTGT-----	TTKTPVTEYPSHEVDIDRTYSAWLPSATIEFK	659	
Qy	586	--DYEMPFASYSRTHRPNTQEMVF--	SOJGDS-----	GVHTALKPERANTWQGFNTYK	636	
Db	660	PVDMFRPYVSYSLRPPTILETFFAGARPGDSAGYEPNQSLRAEKATTYEIGANMSF	719			
Qy	637	KGLLKQDDTLGLKLVGRSRIDNYI-----	HNVYKGMWDLNAGNIPSWSVSTGLAYTIQ	689		
Db	720	DGVLLDDTLRIKMAAFRRVKDYIALGLVTDQVDRTY-----	TSFVNLDTTY---	770		
Qy	690	HRNFKDKVHKGFLELNYDYGRFTWLSYAYOKSTOPTNFSDASEPNNASKEDQLKQG	749			
Db	771	-----	MRGLENGYDARSFWIGSATVLKTEWPEKTQVFSNNTTTTSGE-----	815		
Qy	750	YGLSRVSALPRDYGRLVGRWGLGNKLTGAMRYFKSIRATAEERYIDGTNGNTSNV	809			
Db	816	-----	IYANFGD-----	VAFKM---	KUTLDGMRFFDEKFSGLARLNHVPTQSRILTDE	862
Qy	810	ROLGKRRIKOTETLARQPLIFDYAAEYEPKKNLIFRAEVKNLFDPRYIDPLDAGNDAATQ	869			
Db	863	GNL--REITDPYTTV-----	DLYGSYAFNDKATLRFVAVNLTDRKYIPAASA-----	907		
Qy	870	RYSSFPDPKDKDEVTCAADKTLNCKYKGTSTSVLTFNFAFGRTFLITMSYKF	922			
Db	908	--YT-----	APGRFTIATMNVKF	923		
RESULT 10						
D82437						
TonB receptor-related protein VCA0625 [imported] - Vibrio cholerae (strain N16961 serogroup O1)						
C:Species: Vibrio cholerae						
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001						
C:Accession: D82437						
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.						
Nature 406, 477-483, 2000						
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.						
A:Reference number: A82035; MUID:20406833						
A:Accession: D82437						
A>Status: preliminary						
A:Molecule type: DNA						
A:Residues: 1-784 <HEI>						
A:Cross-references: GB:AE004392; GB:AE003853; NID:g9658031; PIDN:AAF96526.1; GSPDB:GN001						
A:Experimental source: serogroup O1; strain N16961; biotype El Tor						
C:Genetics:						
A:Gene: VCA0625						
A:Map position: 2						
Query Match 6.3%; Score 308; DB 2; Length 784;						
Best Local Similarity 21.2%; Pred. No. 1.4e-11;						
Matches 202; Conservative 115; Mismatches 314; Indels 320; Gaps 46;						
Qy	6	RLKPCIFYLMGVLMYHHSYAEDAGRAGSEAOIQVLEDVHVHAKRVPKDKKVFVDARAVST	65			
Db	36	QLKPL-FTLLPVLLSSVVQAQE-----	NTEQAVDETIVH-----	QGSILTDQR---T	79	

QY	66	RQDIFKSSSELDNIVRSIPGAFQTQDKSSGIVS-----	-----LNIRGDSGFGRVN	111
DB	80	RSOLDK-----	-----VRGIANA-----DIFSGITTSVQSNMHNHEAGALDIGIRGVOGEGRVP	127
QY	112	TWVDGITOTFYSTDAGRAGSSQFGASVDSNFAGLDVVKGS-----	-----FSGSAGINS	164
DB	128	IFIDGSLQ-----	-----STHTSRGYGVSDR---TYIDTDLSSLTNKGATIESSPYASGAVG---	179
QY	165	LAGSANLRTLGVDVVQGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYG		224
DB	180	--GVNATTLIGIKDIIKDDQAGVVLKARANNNHRTPDV-----	-----SGDYSEQG-----YA	227
QY	225	HSRRSVAQNTVRG-----	-----GGGOHIGNFGA--EYLERERKQRYFVQEGGLKFNNSGKWERDF	278
DB	228	LDERGEHSAPFKHGLSLMLGLGYQAESFNTVLAYSKRSGKNHF-----	-----AGK-----	271
QY	279	QRPYWKTKWQKYNDP-----	-----QELQKYIEGHDKSWRENLAPOYDITPIDPSSLKQQSAGN	333
DB	272	-----	-----KGEEYQEPVVGQGEVNTSFESD--SW-----	298
QY	334	LFKLEYDGVFNKYTAQFRDLNTRKIGSRKIIINRNYQFNYSLSNAYANLMLTAAVNSGRQK		393
DB	299	LPKLASD-----	-----TGTAHNADFNRYHHAQKAGEVLMAYWKSSDWEGN-----PYPDGKDR	349
QY	394	YPKGSKFTGWLGLKDFETYNNAKILDLNNTATFRLPRETELOTLTGFNFHN-----	-----EY	447
DB	350	MPQ-----	-----WGL-----GTAKV-----NTYSANYQDPHPWLNL	378
QY	448	GKNRFPPEELGLFFDGDQDNGLYSYLGRFKGDKGLLPKOKSTIVQAGSOYFNTEFYDAA--		506
DB	379	NANFWTEADL-----	-----AQYNGLWA-LG-----TNAEQFYHAYHNDRSG	415
QY	507	-----	-----LKKDIYRLNYSNTNVGRFEGEYTYGYGSDDEFKRAFGENSPTYKHKCNQSC	558
DB	416	LSLTNETLLTOWPVRVNLNYGLAQOONERLSPEEDG---	-----QTRFTKT-----	455
QY	559	GIYEPVLKVKYKRRANHSVSIADFG-----	-----DY-----	587
DB	456	-----	-----VTSRHGKRTAQNLFANADIDYSPLRVQLGNLJHNAKSTDYQTKQOLDYKEKIDLL	510
QY	588	-----	-----FMP-----FASYSRTHRPNIQEMYFS--QIGDSGVHTALKPERANTWQGFNTYK	636
DB	511	SEETAYLTPSTQLFLKSSRTYRMPSLYETTLNSNEVFSYNPNPIKPEQANNVGVQVMA		570
QY	637	KGLLKQDDTLGLKLVGRSRIDNYIHNVYKGMWDLNIPSWSVSTGLAYTIOHRNFKDK		696
DB	571	SNVQLQDDRLNLSVSVFRNSIKDFISG--	-----GRUKATPG--MSEW--QANFTFT-----NY-DK	620
QY	697	VHKHGELELNIDYGRFETNLVAYQKSTOPTNFSDASEPNNASKEDOLKQGVGLSRVS		756
DB	621	LQLSGMELGAHYQYAWLYTHFAATLYSETKCSVQQAQYAESDTC---	-----NSLGFAMGLTPTPR	678
QY	757	ALPRDYGRLEVGTRFWLGNKLTLCGAMRYF--	-----GKSIRATAEERYIDGTNGGNTSNVRQLGKR	815
DB	679	IPPKQNLVNVGTHKFFNDTLDGSKVSVYHSGKS-----	-----NPSDWLAGT-----	720
QY	816	SIKQETTLARQPLI-----	-----FDFYAAEYEPKKNLIFRAEVKNLFDRIIDP	859
DB	721	-----	-----AANPILEIPSDYTDLYSQYELNANTQLFFAINNVTDRYQVRP	763
RESULT 11				
A43335				
transferrin-binding protein 1 - Neisseria gonorrhoeae				
C:Species: Neisseria gonorrhoeae				
C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999				
C:Accession: A43335				
R:Corneilissen, C.N.; Biswas, G.D.; Tsai, J.; Paruchuri, D.K.; Thompson, S.A.; Sparlin				
J. Bacteriol. 174, 5788-5797, 1992				
A:Title: Gonococcal transferrin-binding protein 1 is required for transferrin utiliza				
A:Reference number: A43335; MUID:92394880				
A:Accession: A43335				

RESULT 14
E82443
heme transport protein Huta VCA0576 [imported] - Vibrio cholerae (strain N16961 serogrou
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: E82443
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:title:--DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: E82443
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-698 <HEI>
A:Cross-references: GB:AE004388; GB:AE003853; NID:99657979; PIDN:AAF96478.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0576
A:Map position: 2

Query Match 5.4%; Score 266; DB 2; Length 698;
Best Local Similarity 18.9%; Pred. No. 5.5e-09;
Matches 173; Conservative 123; Mismatches 339; Indels 282; Gaps 33;
QY 23 SYAEDAGRAGSEAQIQVLEVDVHKAKRVPKDKKVFDDARAVSTRQDIEKS-SENLDNIVR 81
DB 24 AYADDYAS-----FDEVVYSTRFLNTQITDTRASVAVINASDIEQMAEDIEGLFK 74
QY 82 SIPCAFTQDDKSGVSLNIRGDSGFRVNTVMDGITQTFYSTDAGRAGGSSQFGASV 141
DB 75 YTPCVTLTNSRGVOCINIRGIEG-NRIKVIDVGAQ---PNOFDSGNSFLNS-RVDI 129
QY 142 DSNFIAGLDVVKGSFSGSAGINSLAGSANLRTILGVDVVGQNTYGLLLKGLTGNTSKG 201
DB 130 DTMVKSVEIVGAASLQSGSDAIGGIVAFETKDPADILKGRNMGGYAKLINSDDKTF 189
QY 202 NAMAAGIARKWLESAGSVGLYGHRSRVAQNYRVGGGGHIGNFGA-----248
DB 190 ESTALANKSDLES-----LVAYTRD-----GQEIQFGSPDOODNANNLLV 233
QY 249 --EYLERRKORYFVQEGGLKFNNSGKWERDQRPYKWKYQKYNPDQLOKQYIEGHDK 306
DB 234 KLOYLNPKHR-----LEFSGNYIRKNNDLEN--LEFSGYKNASCTDETTOYQLGIK 284
QY 307 SW-----RENLAPOYDITPIDPSSL---KQOSAGNLFLKLEYGVNPKYTAQFRDLNTKI 357
DB 285 IWDAEFLADRIITWQDFVVGKEETGITDRTSKNGNIQKDY-----LYSDKGFSFDSQL 339
QY 358 G-SRKIIIRNYQFNYSGLSLNSYANLNLTAAYNSGRQKPKGSKFTGWGLLKDPETYNNAK 416
DB 340 DKSFMYNSNTHYIVYGFSLD-----KDIENTNQEF 370
QY 417 ILDLNNTATRLPRETELOTLTGFNFHNEYKGNRPPEELGLFFDGDQDNGLYSLGRF 476
DB 371 NSIGKKNVIFYIPNASEKR--YGF-FIQDEIAFDNLIVTPGIRFDSFETKPG-----419
QY 477 KDKGILLPKQSTVQPA---GSOYFNTFFDAALK---DIYRLNYSNTVGYRFGGEYT 530
DB 420 -----DTSANPSLNDASEYKK--YSDSALTARLGTGVYKLNQENRL-----457
QY 531 GYYGSDDEFRACFGENSPYKKHCNCSGCIYEPVLYKYGKRRANNSHVSISADFGDYFMP 590
DB 458 -----457
QY 591 FASYSRTHRPNIQEMYFSQIGDSGVHTALKPERANTWQFGENTYKGLLKQDDTLGLKL 650
DB 458 FAQISQGFAPDPQELYS-FGNPAHYVFKP-NPNLEAEDSVSYELGWRYNADSVSNEL 515
QY 651 VGRSRIDNIIHNKWKMDLNGNIPSWSVSSTGLATYIOHRNPK-DKVHKHGFELNLD 709

Db 516 SIFYSYDNFIDS-----QIVSGSEKTRDAVHQHSINIDKATIKGIELSNQFF 562
QY 710 YGRFFTNLSVAYOKSQPTNFSDASESPNNAKEDQLKQYGLSRVSALPRDYGRLEVG 769
Db 563 WDRFMPIVGFSRIAAYATEKDGNGKPLN-----SVSP-----596
QY 770 RWLGNKLTLCGAMRYEFGKSTRATAEBRYIDGTNGGNTSNVROLGKRISIKQETLARQPL- 828
Db 597 -W-----NAVGTGYNDSNNWGTAVNLTYTAKKASEING-DYQPI 636
QY 829 ---IFPFYAAEPKKNLIFRAEVKNLFDRYIDPLDAGNDAATQRIYSSFPDKKDEEVT 885
Db 637 SATVIDVTAYYKPIKDLTLRAGVFNLT-----EYYNWNVDVRLPSE-- 679
QY 886 CNADKTLCLNGKYGGTSK 902
Db 680 -DKDKTQAKRNFGITAK 695
RESULT 15
B64049
outer membrane protein hxuC homolog - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: B64049
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: B64049
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-744 <TIGR>
A:Cross-references: GB:U032696; GB:L42023; NID:gl573057; PIDN:AAC21789.1; PID:gl573065

Query Match 5.4%; Score 266; DB 2; Length 744;
Best Local Similarity 20.5%; Pred. No. 6e-09;
Matches 181; Conservative 132; Mismatches 353; Indels 218; Gaps 39;
QY 14 LMGMVLIHHSYADAGRAGSEAQIQVLEVDVHKAKRVPKDKKVFDDARAVSTRQDIFK-S 72
Db 15 LVGIQL-NVTAKONSSNSRRELLPIIVNTDDSNKLP-----GRSVLKOKNIEQX 65
QY 73 SENLDNIVRSIPCAFTQDDKSSGIVSLNIRGDSGFRVNTVMDGITQTFYSTSTDAGRAG 132
Db 66 ADNAANLILNLPGVNNMAGGFRPGGQTLNMGMDAEDVRVQLDGATKSEK-----116
QY 133 GSSQFCASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTILGVDVVGQNTYGLLLKG 192
Db 117 -YQGSFIFTEPELLRRVTVDKGNYSPOYGNNGGAGTIVKETKDARDFLQENKIGFLK- 174
QY 193 LTGTNSTKGNAMAAIGARKWLESAGSVGLYGHRSRVAQNYRVGGGGHIGNFGAEYLE 252
Db 175 -----YGNNSNNQKTYST-----ALVLQ 193
QY 253 RRKQRYFVQEGGLKFNNS--NSGKWERDQRPYKTKYQKYNPDQLOKQY-----300
Db 194 NEQKNIDL---LLFGSVRNAG---DYRRPDNSKTLFSKNNQKGTGLIKLNWQISPEHLL 245
QY 301 ---IEGHDKSRNENLAPOYDITPIDPSSLKQOSAGNLFLKLEYDVFNKYTAQFRDLNLT 356
Db 246 TLSSVYGIHKGW-EPPFAARDILP-----KPSLSDIMRYGTD-TAWRKRLVYRD-----292
QY 357 IGSRKIIIRNYQFNYSGLSLNSYANLNLTAAYNSGRQK---YPKGSKFTGWGLL--KDFETY 412
Db 293 ---QKDNENTLYKNYLPENNPNWLNLTQFSYKTTQNDMRPKAEASSGLVSLGNQSWITY 349
QY 413 NNAKILDNLNNTATRLPRETELOTLTGFNFHNEYKGNRPPEELGLFFDGDQDNGLYSY 472

Db 350 SDL-TFDINTSTFNI-KTTFVHELLFGLQWLKNT--RNT-----LMYDKSKVRKADYNY 399
Qy 473 LGREFKDGKGLLPKSTIVQAGSQYFNTFYFDAALK-KDIYRLNYSNTVTGVRFGGEYTG 531
Db 400 -GYFQ-----PYMPSGROYTQAFYLDQIQKWNII--FST---GVRY----- 436
Qy 532 YYGSDDEKRAFGENSTYKKHCHQSCGIYEPVLKKGKRRNNHS--VSISADFGDYFM 589
Db 437 -----DHINNIGKNLAL--KYNDISAG-----HDYSQKNYNGWSYVLGLNYDVNHYLS 483
Qy 590 PFASYSRTHRPNIQEMYFSGIGDSGVHTA---LKPERANTWQFGFNTYKKGLKQDDTL 646
Db 484 LFTNFKTWRAPIDEQYETQFQKSSYPATSLNLEKEMINQTRVGGIITLNHLEQENDAF 543
Qy 647 GLKLVGYSRIDNYIH-----NHYGKWDNLNGNI-PSWVSSTGLAYTIOHRNEKDKVHKH 700
Db 544 QFRTTYFNRGKNEIFKTRGVNCGVNAADTNKVCPIIEN-----YRNLPGYVIQ- 594
Qy 701 GFELELNYDYGREFFTNLSYAYOKSTQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPR 760
Db 595 GAELAYYQSTYLEGEITYSVYKGRDT-----SPRN-----PWGKTSTWIAEIPP 640
Qy 761 DYGRLEVGTWMLGNKLTGLGAMRYFGKSIRA-----TAERYIDGTNGGNTSNVRQLGK 814
Db 641 RKATTALGFNVPKYYLTVGWRFAEFVRQDRSPLSGDPKASWSLPSASRGYSLHNL----- 695
Qy 815 RSIQOTETLARQPLIFDYAYAYEPK--KNLIFRAEVKNLFDRRY 856
Db 696 -----FLSWSPAKIKGMNVKITVDNLFNRAY 721

Search completed: July 24, 2002, 08:54:36
Job time: 527 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: July 24, 2002, 08:53:29 ; Search time 20.76 Seconds
(without alignments)
1719.625 Million cell updates/sec
Title: US-09-762-926-4
Perfect score: 4904
Sequence: 1 MRSSRLKPCIFLYMGVLY.....SVLTNFARGRTFLTMSYKF 922
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	2514.5	51.3	913	1	YC17_HAEIN	P45114	haemophilus
2	278	5.7	915	1	TBPI_NEIGO	Q01996	neisseria g
3	269.5	5.5	908	1	TB12_NEIMB	Q06987	neisseria m
4	266	5.4	744	1	HXC1_HAEIN	P44523	haemophilus
5	257	5.2	911	1	TB11_NEIMB	Q09056	neisseria m
6	252	5.1	999	1	HGPB_HAEIN	O87296	haemophilus
7	243	5.0	943	1	LBPA_NEIMB	Q06379	neisseria m
8	242	4.9	912	1	TBPI_HAEIN	P44970	haemophilus
9	241.5	4.9	1013	1	HGBA_HAEIN	O9Kiv2	haemophilus
10	241	4.9	993	1	HGPC_HAEIN	O9Kiv0	haemophilus
11	232.5	4.7	999	1	HGP2_HAEIN	P44809	haemophilus
12	229	4.7	1063	1	HGPI_HAEIN	P44795	haemophilus
13	226.5	4.6	944	1	LBPA_NEIMA	O9Jtk4	neisseria m
14	218.5	4.5	723	1	Y262_HAEIN	P44600	haemophilus
15	211.5	4.3	725	1	HXC2_HAEIN	P45357	haemophilus
16	209.5	4.3	810	1	HPUB_NEIMC	P96949	neisseria m
17	209	4.3	1077	1	HGPA_HAEIN	O9za21	haemophilus
18	205.5	4.2	1084	1	HGP3_HAEIN	P44836	haemophilus
19	204.5	4.2	810	1	HPUB_NEIMA	O9Jwa2	neisseria m
20	192	3.9	1066	1	HGPC_HAEIN	O9x442	haemophilus
21	192	3.9	1067	1	HGBB_HAEIN	O9Kiv1	haemophilus
22	186.5	3.8	1046	1	HHUA_HAEIN	Q48153	haemophilus
23	185	3.8	999	1	HGPA_HAEIN	Q57408	haemophilus
24	181.5	3.7	972	1	HGBB_HAEIU	Q47957	haemophilus
25	180	3.7	614	1	BTUB_ECOLI	P06129	escherichia
26	180	3.7	1567	1	ICEN_XANCT	P18127	xanthomonas
27	179	3.7	676	1	HMUR_YERPE	O56989	yersinia pe
28	175	3.6	687	1	HEMR_YEREN	P31499	yersinia en
29	173.5	3.5	972	1	HGBA_HAEIU	Q47952	haemophilus
30	170.5	3.5	700	1	YNCD_ECOLI	P76115	escherichia
31	169.5	3.5	729	1	PHUE_ECOLI	P16869	escherichia
32	168.5	3.4	774	1	FECA_ECOLI	P13036	escherichia
33	167.5	3.4	1258	1	ICEN_ERWHE	P16239	erwinia her

34 166.5 3.4 1322 1 ICEA_PANAN
35 166 3.4 614 1 BTUB_SALTY
36 164 3.3 710 1 FOXA_YEREN
37 160.5 3.3 815 1 FPVA_PSEAE
38 160 3.3 1148 1 ICEK_PSEEX
39 153 3.1 1200 1 ICEN_PSESY
40 151 3.1 1210 1 ICEN_PSEFL
41 146 3.0 1034 1 ICEN_PANAN
42 145 3.0 782 1 OSTA_HAEIN
43 144 2.9 687 1 VIUA_VIBCH
44 143.5 2.9 809 1 PUPB_PSEPU
45 143.5 2.9 1104 1 COLA_CLOPE

ALIGNMENTS

RESULT 1
YC17_HAEIN STANDARD; PRT; 913 AA.
AC P45114;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable tonB-dependent receptor Hli217 precursor.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischiemann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.",
RL Science 269:496-512(1995).
[2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.,
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32801; AAC22870.1;
DR TIGR: Hli217;
DR InterPro: IPR000531; TonB_boxC.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 913 PROBABLE TONB-DEPENDENT RECEPTOR Hli217.

```
FT SITE 896 913 TONB C-TERMINAL BOX.
SQ SEQUENCE 913 AA; 102768 MW; 01BEFDE5D6AEF617 CRC64;

Query Match 51.3%; Score 2514.5; DB 1; Length 913;
Best Local Similarity 51.3%; Pred. No. 5.5e-151;
Matches 492; Conservative 155; Mismatches 229; Indels 83; Gaps 15;

QY 1 MRSPRLKPICFYLMGVNLYHHSYAEDAGRAGSAQI-QVLEDDVHVAKRVPKDKKVFSTD 59
DQ 1 MKKAIKLNIILGLINTI-----GMTITQAQAEETLQDIDVEKVISNDKKPFTE 50
QY 60 ARAVSTRODIFKSSENLDNIIVSRIPGFTQODKSSGIVSLNIRGDSGFRVNTMVDGITQ 119
DQ 51 AKAKSTREWFKEQTIIDVIRSIPIGFTQODKSGGVVSVNIRGENGLGRVNTMVDGVTD 110
QY 120 TFYSTSTADGRAGSSQFGASVDSNFTAGLDVVGSGSGAGINSLAGSAMLRLTGLVDDV 179
DQ 111 TFYSTALDSGGSGSQFGAAIDFNFTAGVDVKNFSFSGAGINSALAGSANSFRTLGNDV 170
QY 180 VGNNTYGLLLGLTGTNSTKGNAMAAIGARKWLESASGVLYGHSSRSVAQNVRYGGG 239
DQ 171 ITDKPFGIILKMGTSNATSNFTWMAAGRKWLDNGYGVVYGYSGREVSQDIIRI-GG 229
QY 240 GQHTGFGAEYLERRKQRYFYQEGGLKFNSSGKWERDFQRYWTKW--YQKYND---- 293
DQ 230 GERLASLGODILAKEAYFRNAGVIL--NPEGQWTPDLSKKHWCNKPQYKNGDCSY 287
QY 294 -----PELOQY-----IECHKSWRENAPQYDITPIDPS 325
DQ 288 RIGSAAKTRREILOELTNGKKPKDIEKLQKNGDIEETDKSFERN-KQOYSVAPIEPGS 346
QY 326 LKQOSAGLFLKLEVDGKFNKTAQFRDLNTKIGSRKINRNQYQYGLSLNSYANLNTA 385
DQ 347 LQSRSHLLKFEYGDHONLGAQLRLDNKIGSRKINRNQYQYGLSLNSYANLNTA 406
QY 386 AYNGRQKPKGSKFTGWLKLDFTYNNAKILDNNTATFRPRETELOTLGFNFYHN 445
DQ 407 AHNIGTKTIYKGFAGQVADKLITKNVANIVDINNHTFLPKPEIDLTGLGFNFYTN 466
QY 446 EYGNRPPEELGLFDGDDONGLYSY--LGRFKGDKLLPQKTIIVOPAGSQVFNFTYF 503
DQ 467 EYSKRPPEELSLYNDASHDQGLYSHKGRYSKTSKLLPQRSVILLOPSGKQKFTYVF 526
QY 504 DAALKKDIYRLNNTVYGVRFGEYGYGSDDEKFRAGFENSPYTKKHCNQSCGYEP 563
DQ 527 DTALSKGILYLNYSVNTHTAFNCEYGY-----ENTAGQ-----INEP 566
QY 564 VLKYGKRRANHHSVISADGDFYFMPFASYSRTHRPNTIOEYFSGIDSGVHTALKPE 623
DQ 567 ILKSHGKKAHNSATLSAELSDYFMPFTYSTRHMPNIQEMPFQSVNAGVNTALKPE 626
QY 624 RANTWQGFNTYKGLLKQDDTLGLKLVGRYSRIDNVIHNVYKRWDLNIGNISWSSTG 683
DQ 627 QSDTYQGFNTYKGLTQDDVLGVLGVRSFKTKNIHNVYGVW--RDMPTWAEISNG 684
QY 684 LAYTIQHRNFKDKVHKHGFLELYDGRFTNLISYAKSTOPTNFSDAESPNNAKSE 743
DQ 685 FKYTIAHQNKPIYKKSVELEINYDMGRFANVSAYOFTNPTNADASPNNASOE 744
QY 744 DOLQOGGLSRVSLPRDYGRLEVTGRLWKNKLTGLGAMRYFKSIRATAERYIDGTNG 803
DQ 745 DILQOGGLSRVSLPRDYGRLEVTGRLWKNKLTGLGAMRYFKSIRATAERYIDGTNG 803
QY 804 GNTSNRQLGKRSIKOTETLAPLIDFVAAYEPKKNLIFRAEVKNLFDPRRYIDPLDAG 863
DQ 804 FKWTLRNENYAVKKTEDIKQPIILDLHLVSYEPIDKLIKAQVQNLDRYVDPDLAG 863
QY 864 NDAATQRYYSFDPKDKDEEVTNADKTLGNGKYGTGSKVSLNFARGRITLITMSYKF 922
DQ 864 NDAASQRYYSLL-----NNSIECAQDSSAC-----GSGDKTVLYNFARGRITLITMSYKF 913
```



```
QY 290 KYNDPQELQKYEIHKD-----SWRENLAPOYDITPIDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KDDVGEDKQVTVSTDYTGPNRFLA--DPLSYESRSMWLFRCFRFENKRHY 326
QY 340 DGVENKYTAQ---PRDLNKTIGSKRII-----NNYQFNGLSLNSYANLNL 383
Db 327 IGGILERTQOTFDTRMTVPAFLTKAVFDANOKAQSIRGNGKYAGNH-----KYGL-- 379
QY 384 TAAYNSGRQYPKGSKFTGWLKDKFTYNNAKILDNNATFRLPRETELOTTGLGNFYF 443
Db 380 ---FTSGENNAVPVGAEE--GTGVFYD-----ETHTKSYGLEYY 413
QY 444 HNEYGKRFPEELGLFFD--GPDODNGL-----YSYLGRFKGDK 480
Db 414 YTNADKTDWADYARLSYDROGIGLHNFHQTHCSADGSKYCRPSADKPFY---YKSDR 470
QY 481 GLLPKQSTIVOPAGSOYFNTFYDAALKDIIYRLNYSTNTVGY--RFGG-----EYTGYYG 534
Db 471 VIYGESHKLLQAAPKKSFDATAI-----RHNLNVN--LGYDRFGSNLRHODY--YYQ 518
QY 535 SDDFEKRAFGENSPYKHKHNCQSGIYEPVLKYGKKRA-----NNHVSYSISAD----- 583
Db 519 SAN---RAYSLKTPP-----QNNGKKTSPNGREKNPYWVSIGRGNVVT 559
QY 584 ---FGD-----YMPFA-----SYSRTH----- 598
Db 560 QICLFGNTYDCTPRINSKSYTAARVDNVLGRWADVAGLRYDRSTHSDGCVSTG 619
QY 599 -----RMPNIQEMYFSQIGDSGVHTALKPERANTWQ 629
Db 620 THRTLSNAGIVLKPADWLDITYRTSTGFLRPSFAEMYWRSGDKIKAVKIDPEKSPNKE 679
QY 630 FG-----FNTYKGLLKODDILGLKLGYRSIDN-----YIH--- 662
Db 680 AGIVFGDGFNGLEASWNNAYRDLIVR-----GYEAQIKDGEQVKGPNPAYLNAQS 730
QY 663 -----NVYGR--W---WDLNGNIP--SWVSTGLAYT--IQHRNFKDKVHKHGFLELYND 709
Db 731 ARITGINILKIDNGVWD---KLPEGWYST--FAYNRVRVRDIKKRADRTDIOGSHL--FD 784
QY 710 YGREFTNLVYAYQKSTQPTNFSDAESPNNASKEDQLKOGYLSRVSAALPRDYGRLEVGT 769
Db 785 -----AIQPSRYV-----VSGGYDQPEGKGVNGMLT-----YSKAKEIT 819
QY 770 RWLGNKLTGGMARYFKSKSRATAERYIDTNGNTSNVRLKGRKIKOTETILAROPLI 829
Db 820 ELLGSRALLNG-----NSRNTKATARTREPWYI 847
QY 830 DFDEAYAEPPKKNLIFRAEVKNLFDRIYI 857
Db 848 VDVSGYTVVKKHFTLRAGVYNLLNHRIV 875

RESULT 3
TB12_NEIMB
ID_TB12_NEIMB STANDARD; PRT; 908 AA.
AC Q06987;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-42.
RC STRAIN=CCUG 37603 / B16B6 / SEROGROUP B / SEROTYPE 2A;
RX MEDLINE=93345825; PubMed=8344530;
RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
RA Jacobs E., Schryvers A.B.;
RT "Cloning and characterization of Neisseria meningitidis genes
RT encoding the transferrin-binding proteins Tbp1 and Tbp2.";
```

```
RL Gene 130:73-80(1993).
RN [2]
RP SEQUENCE OF 25-45.
RC STRAIN=CCUG 37603 / B16B6 / SEROGROUP B / SEROTYPE 2A;
RX MEDLINE=93307625; PubMed=8319886;
RA Griffiths E., Stevenson P., Byfield P., Ala'Aldeen D.A.,
RA Barriello S.P., Holland J., Parsons T., Williams P.;
RT "Antigenic relationships of transferrin-binding proteins from
RT Neisseria meningitidis, N. gonorrhoeae and Haemophilus influenzae:
RT cross-reactivity of antibodies to NH2-terminal peptides.";
RL FEMS Microbiol. Lett. 109:85-91(1993).
CC -1- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY IRON STARVATION.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z15129; CAA78831.1; -.
DR PIR; JN0819; JN0819.
DR PIR; S33154; S33154.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 908 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 35 45 TONB BOX.
FT SITE 891 908 TONB C-TERMINAL BOX.
FT SEQUENCE 908 AA; 101583 MW; FE2FF4974CAC3C31 CRC64;
QY 5 FRLKPICFYLMGVMLYHHSYAEDACRAGSEAOIOVLDEHVHVKAK--RVPKDKKVFETARA 62
Db 7 FRLNILCLSLMTAL---PVYAEVY--QAEQAQEKQLDTIQVAKKQKTRRDNEVYGLK 61
QY 63 VSTRQDIFKSSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGRVNTWVGIT 118
Db 62 VKSSDTLSKQOVLNIRDLTRYPDGIADVVEQGRGASSG---YSIRG-MDKNVRVSLTVDGVS 117
QY 119 Q-TEYSTSTDAQ--RAGGSSQFGASVDSNFTAGLDVVVKVSGSAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGGTTRTAGSSGAINIEYEVNKAIVEISKGSNSSEYNGCALAGSAVFQTKT 177
QY 176 VDDVVGNNYVGLLLK--GLGTNTSTKGNAMAAICARKWLESGASGVLYGHRSRVAQNY 234
Db 178 AADIIGEGKQWGIQSKTAYSGKDHALTQSLALAGR---SGGAELLITTKRRGREIHAAH 233
QY 235 RVGGGGQHIGNFGAEYLERKKQ-----RYFVQE---GGLKFNSN-----SGKWERDF 278
Db 234 KDAQG--VOSFNRLVDEDEKKEGSGYRYFIVEEECHNGYAAACKNKLKEDASVKYDER-- 289
QY 279 QRPYWKTKWKYKNDPQE-LQKYIEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKL 337
Db 290 -----KTVSTQDITGSRNRLANPLEYGSQSWL--FRGW-----HLDNR 326
QY 338 EYDGVFNKYTAQFRDLNLTIGSRKRIINRNYOFNYGLSINLYANLNTAAYNSGRQYPK- 396
Db 327 HYVGAVLERTQOTEDTRDMTVPAYFTSEDYVPGSLKGLGKY-----SGDKNAERL 376
QY 397 -----GSKFTGWGLLKDFEYNNNAKILDNLTNTATFRLPRETELOTTGLGFNFHNEYGNR 451
```

Query Match 22.5% Score 269.5; DB 1; Length 908;
Best Local Similarity 21.0%; Pred. No. 1.7e-09;
Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;

```
Db 377 FVQEGSTLOGIG-----YGTGVFYD-----ERHTKNRYGVVEVYHNAKDT 418
QY 452 FPEELGLEFD--GPDQDNGL-----YSLGRFKGDKGLLPQKST 488
Db 419 WADYARLSYDROGIDLDNRLOQTHCSHDGSKNCRPGDKKPYSF---YKSDRMIIYESRN 475
QY 489 IVQAGSGYFNT-----FYFDAAL----- 507
Db 476 LFOAVFKAFDIAKTRHNLINLGYDREKSQLSHSDYDLQNAVQAYDLITPKKPPFNGS 535
QY 508 KKDYLRLNYSNTWG---YFEGGE-YT-----GYGSDDEPKRAFGENSPYTKK 552
Db 536 KDNRYRVSIGTKTNTVSPICRFGNNTYTDCPTRNIGNGY---AAVQDNVRLGR 587
QY 553 HCNOSCGI-YEPVLKKYCKKRNHNSVS-----ISADFGDYFMPFASYSRTH-----R 599
Db 588 WADVGAGIRYD-----YKSTHSEDKSVSTGTHRNLNWSNAGVVLKPFPTWMDLTYRASTGER 642
QY 600 MPNIQEMYFSQIGSGVHTALKPERANTWQFG-----FNTYKKGLLKQDDT 645
Db 643 LPSFAEMYGRAGESLTKLDLKPESFNREAGIVFKGDFGNLEASYFNNAVYRDLI----- 697
QY 646 LGLKLGVRSIDN-----YIH-----NVYK--WMDLNGNIPSWSTGL 684
Db 698 ----AFGYETRTONGOTSASGDPGRYNAQNAARIAGINILGKIDWHGVWGLPDGLYST-L 752
QY 685 AY-----TIQHREKDKVHKHGFLELNDY--GREFTNLSVAYOKSTQPT 728
Db 753 AYNRKVKDADIRADRTFVTSYLFDAQVPSRYVLGLGYDHPDGLWGTMTYK----- 807
QY 729 NFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWGLNKLTLGGAMRYFGKS 788
Db 808 -----AKSVDEL-----LGSQALLNG----- 823
QY 789 IRATAERYIDGTNGGNTSNVROLCKRSIKOTETLARQPLIFDYAYAEPKKNLIFRAEV 848
Db 824 -----NANAKKAASRR-----TRPWYVTDVSGYNTKTKHLLTRAGV 859
QY 849 KNLFDRRYI 857
Db 860 YNLLNRYV 868

RESULT 4
HCX1_HAEIN STANDARD; PRT; 744 AA.
AC P44523;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin utilization protein C precursor.
GN HKUC OR H10113.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD";
RL Science 269:496-512(1995).
```

```
CC CC FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
CC CC CONCENTRATIONS (BY SIMILARITY).
CC CC SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC CC SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC CC SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE B OF
CC CC H.INFLUENZAE.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL: U32696; AAC21789.1; -.
CC CC TIGR: H10113; -.
CC CC InterPro: IPR000531; TonB_boxC.
CC CC Pfam: PF00593; TonB_boxC; 1.
CC CC PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC CC PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
CC CC Outer membrane; Transport; TonB box; Signal; Receptor;
CC CC Complete proteome.
CC CC
CC CC SIGNAL 1 24 POTENTIAL.
CC CC FT CHAIN 25 744 HEME/HMOPEXIN UTILIZATION PROTEIN C.
CC CC SQ SEQUENCE 744 AA; 85043 MW; 74D94F72E41AEC31 CRC64;
CC CC
CC CC Query Match 5.4%; Score 266; DB 1; Length 744;
CC CC Best Local Similarity 20.5%; Pred. No. 2.2e-09;
CC CC Matches 181; Conservative 132; Mismatches 353; Indels 218; Gaps 39;
CC CC
CC CC YQ 14 LMGMVLYHHSYAEADAGRAGSEAOIQVLEVDVHVRAKRPDKKVFETDARVSTRODIFK-S 72
CC CC Db 15 LVGIQL-NVTAQKSNSSNREELLPIIVTNDSDSNKLP-----GRSVLKQKNIQSOXQ 65
CC CC YQ 73 SENLDNIVRSIPGAFTQODKSSGIVSLNIRGDSGFRVNTWVGITOTFTYSTSTADAGAG 132
CC CC Db 66 ADNAANLINILPGVNMAGFRPGQGTINMGMDAEDVRVQLDGCATKSFK----- 116
CC CC YQ 133 GSSQFGASVDSNFTAGLDVVVKGSFSGAGINSLAGSANLRTGLVDVVQNNYVGLLKG 192
CC CC Db 117 -YQGSIFIEPELLRRVTVDKGNYSPOYGNGGFGAGTVFKTKDARDPQENKIGGFLK- 174
CC CC YQ 193 LTGTNSTKGNMAAIGARKWLESASVGVLVYHSHRSRVAQNYRVGGGQHGIFGAEYLE 252
CC CC Db 175 -----YGNSSNNNQKTYST-----ALVLQ 193
CC CC YQ 253 RRQRYFVQEGGLKFNFS--NSGKWERDFQRPYWKTKWYQYNDPQELQY----- 300
CC CC Db 194 NEQKNIDL----LLFGSVRNAG---DYKRPNDSKILFSKNNQKTGLIKLNWQISPEHLL 245
CC CC YQ 301 ----IEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEVDGVFNKYTAQFRLDNTK 356
CC CC Db 246 TLSVYGIHKGW-EPPAAKRDLIP-----KPSLSIDIMRYGTD-IAWKRLVYRD---- 292
CC CC YQ 357 IGSRKIIINRYQFNYSGLNSYANLNLTAAAYNSGRQK--YPKGSKFTGWGL--KDFET 412
CC CC Db 293 ---QKDNQYTLKYNLPENNPNWLNSTQFSYSKTTQNDMRPKKEASSGLVSLGNSQSWIT 349
CC CC YQ 413 NNKILDLNNTATFRLPRETELQTTLCGFNFHNEYGKNRFPPEELGLFFDQDQNGLYSY 472
CC CC Db 350 SDL-TFIDINNTTNI-KTTVHELLFGLWLKNT--RNT-----LMYDKSKVRKADYNY 399
CC CC YQ 473 LGRFKGDKLLPQKSTIVQAGSQYFNTFYFDAALK-KDIYRLNYSNTVGYRFGGEVYG 531
CC CC Db 400 -GTFQ-----PYMPSGRQYTOAFYLQDQIKWNII----FST---GVRY----- 436
CC CC YQ 532 YGSDDEKFRAGFGENSPYTKKHNCQSGIYEPVLKKGKRRNNHS--VSIADFGDYFM 589
CC CC Db 437 ----DHINNIGQKNLAL--KYNDISAG-----HDYSQKNYNGWSYLLGLNVDVNHLS 483
CC CC YQ 590 PFASYSRTHRMPIQEMYFSQIGSDSGVHTA---LKPERANTWQFGFNTYKGLLKQDDTL 646
```

```

Db 484 LFTNFSKTRAPVDEQYETQFOSSVPATSLNLEKEMINQTRVGGIITLNHLPQENDAF 543
QY 647 GLKLVGRSIRDNVYH-----NVYKGMWDLNGNI-PSWVSSTGLAYTIOHRNFKDKVHKH 700
Db 544 QFRTTYFNRGKNEIFKTRGVNCVGNAAATNNKCPKIEN-----YRNLPGYVIQ- 594
QY 701 GFELELYNDYGRFTNLISYAYQSTQTFNFSDASESNNASKEDOLKQGLSRVSLPR 760
Db 595 GAELIAYQSTYLFGETYSYVGRKDT-----SPRN-----PMWKTSTWIAEIPP 640
QY 761 DYGRLEVGTRWLNKGLTGLGAMRVFGKSIRA-----TAERYIDTNGNGNTSNVRQLGK 814
Db 641 RKATTALGFNVPKYLLVGNRAEVRQDRSPLSGDPKASSWSLPASRGISLHNL----- 695
QY 815 RSIKQETLARQPLIFDYAYEPEK---KNLIFRAEYKNLFDRRY 856
Db 696 -----FLSWSPAKIKGMNVKITVDNLFNRAY 721

RESULT 5
TBII_NEIMB STANDARD; PRT; 911 AA.
AC Q09056;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPI.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CCUG 37608 / M982 / SEROGROUP B / SEROTYPE 9;
RX MDLINE=Q3345825; PubMed=8344530;
RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
  Jacobs E., Schryvers A.B.;
RT "Cloning and characterization of Neisseria meningitidis genes
  encoding the transferrin-binding proteins Tbp1 and Tbp2.";
RL Gene 130:73-80(1993).
CC -1- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
  TRANSFERRIN UTILIZATION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY IRON STARVATION.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15130; CAA78833.1; -.
DR PIR; JN0821; JN0821.
DR PIR; S33156; S33156.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 911 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB_BOX.
FT SITE 894 911 TONB_C-TERMINAL_BOX.
FT SEQUENCE 911 AA; 101631 MW; 99283ABAE0B773E6 CRC64;

```

Query Match 5.2%; Score 257; DB 1; Length 911;
 Best Local Similarity 21.5%; Pred. No. 11e-08;
 Matches 218; Conservative 131; Mismatches 356; Indels 308; Gaps 55;

```

RESULT 6
HGPB_HAEIN STANDARD; PRT; 999 AA.
ID HGPB_HAEIN

```

```

QY 5 FRKPICFYLMGVMLYHHSYAEDAGRAGSEAAQIOVLIEDVHVAK--RVPKDKKYFTDARA 62
Db 7 FRLNILSLMTAL---PAYAENV-QAG-QAEKQOLDTIOVKAQKQKTRDRNEVTGLGKL 61
QY 63 VSTRQDIFKSSSE-NLDNIVRSIPG-AFTQODK---SSGIIVSLNIRGDSGFGFVNTWVDGIT 118
Db 62 VKTADTLSKQOVLDIRLTDYDPGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVVDGLA 117
QY 119 Q-TFYSTSTDAG--RAGGSSQFGASVDNFIAGLDVYVKGFSFGSAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGCTRTAGSSGAINIEYENVKAVEISKSGNSVEQGSAGALAGSVAFTQKT 177
QY 176 VDDVQGNNTYGLLLKLTGTNSTKGNMAAIGARKWLESGASVGVIYGHHSRVSVAQNYR 235
Db 178 ADDVIGEGROWGIQSK-----TAYSGKNRGLTQSLA 208
QY 236 VGGGQGHIGNFGAEYLERRKQRYFVQEGGLKFNSNSCKWERDFOR--PYWKTWKYQYND 293
Db 209 LAG---RIG-GAEALLIHTGR---RAGEIRAHADAGRGVQSFNRLVPVEDSSYAYFIV 260
QY 294 PQELQ-----KYIEGHDK---SWRENLAPOYDITPIDPSLKKQOS-----AGN 333
Db 261 EDECEGKNYETCKSKPKKDVGVKDERQTVSTRDYTGPNRFLA--DPLSYESRSWLFPRGF 318
QY 334 LF--KLEYDGVFNKYTAQ---FRDLNTYKIGSRKIINRYQFNGLSLNSYANLNLTAAYN 388
Db 319 RFENKRHYIGGILEHTQOTFTDRTMTVPAFLTKAV-----FDANSKQAGSLPGNGKY- 370
QY 389 SGROKYPKSGKFTGW---GLLKDFETYNNAKILDLNNTATFRLPRETELOITLGFNYEHNE 446
Db 371 AGNHKY--GGLTNGENGALVGAE-YGTGVFYD-----ETHTKRGYLEYVYTN 416
QY 447 YGKNRPPEELGFDF--GPDQDNGL-----YSYLGRFKGDKGLL 483
Db 417 ADKDTWADYARLSYDROGIGLDNHFOQTHCSADGSKYCRPSADKPFYSY---YKSDRVY 473
QY 484 PQKSTIVQAPAGSQYFNTF-----YFDAALK-KDIYRLN-----YSTWT----- 520
Db 474 GESHRLLQAAAFKKSFDATAIRHNLSVNLGDFRDFSRLRHODYYOYOHANRAYSSKTPPKTA 533
QY 521 -----VGYRFGGEYTG---YYGSD---DEFKATGENSEPTYKKHCNQCQGIYEP 563
Db 534 NPNGDKSKPYVWSIGGNNVVTGQICLFGNNYTDCTPRSNGKSYAAVRDNLVLRWAD 593
QY 564 V----LKKYKGRANHSVS-----ISADFGYFMFASFYSRTH-----RMPNIQEMY 607
Db 594 VGAGLRDYRSTHSDGCVSTGTHRTLSSWAGIVLKPADWLDLTYRTSTGFRLPSPAEMY 653
QY 608 FSQIGDSGVHTALKPERANTWQFG-----FNTYKGLLKQDDTLGLKLVGY 653
Db 654 GWRSGVQSKAVKIDPEKSPNKEAGIVFKGDFGNLEASWFNNAYRDLIVR-----GY 704
QY 654 RSRIDN-----YIH-----NVYK--W---WDLNGNIP-SWVSSTGLAYTI 688
Db 705 EAQIKNGKEAKGDPAYLNAQASRITGINLIGKIDWNGVWD---KLPEGWST--FAYNR 759
QY 689 QH-RNEFKDKVHKHFELELNYDY---GRFTNLISYAYOKSTQPTNFSDASESPPNASKED 744
Db 760 VHVRIKIKRADRTDIQSHL-FDAIQPSRYVVGLGY-----D 794
QY 745 OLKQGYGLSRVSLPRDYGRLEVTGRWLGNKLTGLGAMRYFGKSIRATAERYIDGTNGG 804
Db 795 QPECKWGVNGMLT-----YSRAKEITELLSRALLNG----- 826
QY 805 NTSNVRQLGKRSIKQETLARQPLIFDYAYEPEKKNLIFRAEYKNLFDRRYI 857
Db 827 -----NSRNTKATARTRPWYIVDVSGYITIKKHTLRAGVYNLLNRYV 871

```



```

Db 772 HNDWGFSTSLFKTKYKFNIDILFKGKD-----FPLVSGGSLPFLSLYNINRDNASLK 827
QY 701 GFELELNYDGR-----FFTNLSYAYQKSTOPTNFSDASESPNNASKEDQLKQCYGLS 753
Db 828 GIEINSKVLGKNAKEMDGFENLSKYKYQKGRNGNI-----PMNAIQPRTWVYGLGYD 881
QY 754 RVSALPRDYGRLVGRVWLGNKLTILGAMRYFGKSTRATAERYIDGTNGGNTSNVRQLG 813
Db 882 HPN-----HKFGDFYTT-----HVASKNPEDYDIYAKD 911
QY 814 KRSIKOTET-----LARQPLIFDYAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAAATQ 869
Db 912 K----KOTDSIKWRSKSYTILDIGYVQPKNTLTIRAGYNLTNRKIYT-----WDSA-- 961
QY 870 RYSSFDPKDKDEVTNCADKTLGCKYGYGTSKVLTFNFAGRGTFLTITMSYKF 922
Db 962 RSRISGTSNVIEQ-----STGLGINRFAVGRNYKMSVQFEF 999

RESULT 7
LBPA_NEIMB
ID LBPA_NEIMB STANDARD; PRT; 943 AA.
AC Q06379; Q9JVK5;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactoferrin binding protein A precursor (Iron-regulated outer membrane protein A).
DE LBPA OR IRGA OR NMB1540.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BNVC / SEROGROUP B;
RX MEDLINE=94011384; PubMed=8406871;
RA Petterson A., van der Ley P., Poolman J.T., Tommassen J.;
RT "Molecular characterization of the 98-kilodalton iron-regulated outer membrane immun. of Neisseria meningitidis.";
RL Infect. Immun. 61:4724-4733(1993).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.";
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: BY IRON STARVATION.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; X69214; CAA49148.1;
CC EMBL; AE002504; AAF41895.1;
CC TIGR; NMB1540;
CC InterPro; IPR000531; TonB_boxC.

```

```

DR PFam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 943 LACTOFERRIN BINDING PROTEIN A.
FT SITE 826 943 TONB C-TERMINAL BOX.
FT CONFLICT 8 8 P -> Q (IN REF. 1).
FT CONFLICT 16 18 IAT -> VAA (IN REF. 1).
FT CONFLICT 22 22 A -> S (IN REF. 1).
FT CONFLICT 26 31 QAGGAT -> NPETAA (IN REF. 1).
FT CONFLICT 43 43 I -> V (IN REF. 1).
FT CONFLICT 64 64 V -> A (IN REF. 1).
FT CONFLICT 233 233 R -> H (IN REF. 1).
FT CONFLICT 243 243 E -> A (IN REF. 1).
FT CONFLICT 247 247 D -> N (IN REF. 1).
FT CONFLICT 257 269 DIKRTREFFSV -> GIKKPSGEGEYFLA (IN REF. 1).
FT CONFLICT 273 275 RES -> SEL (IN REF. 1).
FT CONFLICT 281 281 L -> V (IN REF. 1).
FT CONFLICT 284 286 YGK -> NGN (IN REF. 1).
FT CONFLICT 313 313 Q -> M (IN REF. 1).
FT CONFLICT 389 389 E -> K (IN REF. 1).
FT CONFLICT 420 423 KNLV -> OKLI (IN REF. 1).
FT CONFLICT 455 455 A -> K (IN REF. 1).
FT CONFLICT 546 546 K -> N (IN REF. 1).
FT CONFLICT 564 572 STGFEDNNQ -> YSDYTDKG (IN REF. 1).
FT CONFLICT 658 658 L -> V (IN REF. 1).
FT CONFLICT 667 667 V -> L (IN REF. 1).
FT CONFLICT 943 AA; 105680 MW; BD569ECACFC01A84 CRC64;
SQ SEQUENCE

```

Query Match 5.0%; Score 243; DB 1; Length 943;

Best Local Similarity 20.3%; Pred. No. 8.5e-08;

Matches 224; Conservative 150; Mismatches 375; Indels 352; Gaps 58;

QY 9 PICFYLMGVMLYHHSVAEDAGRAGSE-AQIQVLVEDVHVAKRVPKKKVTVDARAVSTRQ 67

Db 8 PLTLTALATATAPPAYAAQAGGATPDAAQTQSLKEITVTRAAKVGRRSKTEGL-----G 61

QY 68 DIPKSSENLD-----NIVRSIPGAFTQODKSSGIVSLNIRGDSGFGVRNVMVDGITO 119

Db 62 KIVKTSETLNKEQVLGIRDLTRYDPGVAVVVEQNGASGGYSIRGVOK-NRVAVSDVDAQ 120

QY 120 ----TFYSTSTDAGRAGSSQFGASVDNFIAGLDVYKGSFGSAGINSLAGSANLRTLG 175

Db 121 IQAFTVQGSLSGYGGRGGGAIN-EIEYENISVTEIDKGAGSDHSGALGGAVAFRTKE 179

QY 176 VDDVQGNNTYGLLLKGLTGNTSTKGNMAAIGA-----RKWLES-----G 216

Db 180 AADLISDGKSWGIOAK--TAYGSKNRQFMKSLGAGFSKDGWEGLLIINTERQGRETRPHGD 237

QY 217 ASVGVLVGHSHRRSVAQRYNRVGGGGQHNGFAEYLLERRKOR--YFVQEGGLKFNSNSGKW 274

Db 238 IADGVEYIDR-----LDAPQTYDIKRTREFFFSVEG----- 271

QY 275 ERDFQRPYMKTKWYKYNPDQELQKYEIGHDKSWRENLAPODYITIDPSSLKQOOSAGNL 334

Db 272 ERE-SKPVAKLAGYGYKLNQ-LNRWYKER-----IEQNPLSAEEE----- 311

QY 335 FKLEYDGVFNKYTAQPRDLNLTGKSRKINRNQYFNGLSLNSYAN----LNLTAAVNSG 390

Db 312 -----AQVREAQAR-----HENLSAQAYTGGGRLTLPDPMYRSG 345

QY 391 RQKYPKGSKFTG---WGLLKDFETYNNAKILDLNNTATFLPRETELQTLGFNFYHNE 446

Db 346 SWLAKLGYRFGGHHYGVGFED--TKQDYDIRM-----TEKQ-----YIGTD 386

QY 447 YGKNRFEELGLFFDGDQDNGLY--SYLGRFRKGDKLLPQKSTIYOPAGSOYFNTFYFD 504

Db 387 EAE-KFRDKSGV-YDGDGDFRDLGYFVFNIEWKGNL-----VRGIGLKYSRTKFID 437

```
QY 505 AALKKDIYRLNVTNTVG---YRFGGEVYGYGSDDEF-----KRAFGENSPYKKHQNQC 558
D 438 E-----HRRRMLGLLYRENEAGYSDNWADKAVLSPDKQGVATDNNTLK-----LNC 484
QY 559 GIYEPVLKYGKRRANHSSYSISADGDPMPFASYSRTHRMENIQEMFYSQ-----610
D 485 AVY-PAVDKSCRASADK---PYSYDSDRF---HYREQH---NVLNASPEKSLKNNKWK 533
QY 611 -----IGDSGVHTALKPER-----ANTWQFGENTYKGLLKQDDTLGLKLGVY---653
D 534 HHLTLGFGYDASKAISRPEQLSHNAARISESTGFENNDQKYLGLGKPEVGVSGVCIET 593
QY 654 -RSR--IDNVIH--NVY-----GKWD--LNGNIP-----SW 678
D 594 LRSKCVPRKINGSNIHISLDRFSICKYDFDSLGGYDRKNETTSEELVRSGRYVDRSW 653
QY 679 VSTGLA-----YTIQHR---NFKDKVHKHGFLELNDY-----710
D 654 --NSGLFKPNRHFSYSIRASSGFRTPSFQELFGIDYHDPKGWOPALKSEKAANREI 711
QY 711 -----GRF-FTNLSYAYQKSTOPTNFSF-ASESPNNASKEDQL-----KQGYGLSRVS 756
D 712 GLQWKDGFLEISSFRNRYTDMIAVADHKTLPNQAGQLTEIDIRDYNAQNNWSLQGVN 771
QY 757 ALPR-----DYGLEVG--TRWLGKLTGGAMRYFGKSIK-----ATAERYIDG----800
D 772 ILGKIDWNGVYGLPEGLYTTLAYNRKPKSVSNRPGLSLRSYALDVAOPRSYVLGFGYD 831
QY 801 -----TNGGNTSNVROL--GKRSIKQETETLARQPLIFDFAAYEPKKNLIER 845
D 832 QPKGKGANLMLYSKGNKPNDELAAYLAGQKRYSTKRASSWSWADVSAYLNKLLKLTLR 891
QY 846 AEVKNLFDRIYDPLDAGNDAATQRYSSRDPKDKDEEVTCTNADKTLGNCYKGTGSKSVL 905
D 892 AAIYNICNRYV-----TWESLRQTAESTAN-----RHGSDS-----923
QY 906 TNFAR-----GRTELTWYSYKF 922
D 924 -NYGRYAAPGRNFSALAMKF 943

RESULT 8
TBPL_HABIN STANDARD; PRT; 912 AA.
AC P44970;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable transferrin-binding protein 1 precursor.
GN TBPA OR TBPI OR HI0994.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald J.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD";
RL Science 269:496-512(1995).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
```

```
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32780; AAC22656.1; -
CC TIGR: HI0994;
CC DR InterPro: IPR000531; TonB_boxC.
CC DR Pfam: PF00593; TonB_boxC; 1.
CC DR PROSITE: PS00430; TONB_DEPENDENT_REC.1; 1.
CC DR PROSITE: PS01156; TONB_DEPENDENT_REC.2; 1.
CC KW Outer membrane; receptor; Signal; TonB box; Complete proteome.
CC FT SIGNAL 1 23 POTENTIAL.
CC FT CHAIN 24 912 PROBABLE TRANSFERRIN-BINDING PROTEIN 1.
CC FT SITE 50 57 TONB BOX.
CC FT SITE 895 912 TONB C-TERMINAL BOX.
CC SQ SEQUENCE 912 AA; 103233 MW; 294C08991A652CC2 CRC64;

Query Match 4.9%; Score 242; DB 1; Length 912;
Best Local Similarity 20.2%; Pred. No. 9.4e-08;
Matches 213; Conservative 136; Mismatches 419; Indels 288; Gaps 49;

QY 5 FRLKPICFVLMG--VMLYHHSVAEDAGRASGEAQIQVLED-----VHVAKRV--PKDKK 55
D 7 FRLSIISCLLISCYKVAETQSIKDTKEAISSEVDYDTOSTDESELETISVTAEKYRDRKONE 66
QY 56 VETDARAVSTRODIFKSSB-NLDNIIVRSIPGAFTQDKSSGIVSLNTRGDSGFRVNTMV 114
D 67 VTGLGLIKITSEISREQVNLNRLTRYDPGISVVGRCASSGYISRG-MDRNRVALLV 125
QY 115 DGIOTQ-FYSTSTDAGRAGGSSQFGA--SVDSNFIAGLDVYKVSFGSGAGINSLAGSANL 171
D 126 DGLPQTSYVWQSPVARSYSGTGAINEYENKAVEISKGSSEYNGALAGSVTF 185
QY 172 RTLGVDDVVQGNNTYGLLLKGLTGTNSTKG--NAMAAGARKWLESGASGVLYGHSRRS 229
D 186 QSKSAADILEGDKSWGIGQTKN-AYSSKNKGFTHSLAVAG-----KGGFEGELAIYTORNS 239
QY 230 VAQNYRVGG--GGQHIGNFGAEYLERRKQRYFVQE-----262
D 240 IETQVIKDKALGVQSYNRLIAK--PENQSAFYVMEDECPCPKYDECIPSAKPAILSTKKE 297
QY 263 -----GGLKPNSSNGKWERDFQRYWTKWYQYNDPQELQKYIEGHDKSWRENLA 314
D 298 TVSVSDYTGANRIKPNPMKYE-----SQSWFLRGYHFSQHYIGGIFEFYEQ---345
QY 315 QYDITPID-PSSLKQOQAGNLF-----KLEY-----DGVFNKYTA--QPRDLNTK--356
D 346 KFDIRDMTFPAYLSPTKPGDLANRPYPKODYGAYQHIEDRGVGVYASGLYDFDEHRRKOR 405
QY 357 -----IGSRKIINRYQFNYSGLNSYANLNLTAAYNSGRQKYPKGSKFTG 402
D 406 VGIEYIYENKNAKAGIDKAVLSANQQ---NIILDSYR-----HTHCSLYPNPSKNCR 455
QY 403 WGLLKDFEYNNNAKILDLNNTATFRLPRETELQT-----TLGFNRYF-----HNEYG 448
D 456 PTLDRPYSYVSRDNYVKEKHNMLQNLLEKKIQNNWLTHQIVFNGLGDFDDFTSALQHKDYL 515
QY 449 KNRFEPEELGLFED--GPDQDNGL--YSYLGRPKGDKLLPQKSTIVQPSAGSYFNTFYFD 504
D 516 TRRVITATADSISDKTGKTRNGLREYPYL-----YPKPKPYF-----552
QY 505 AALKKDIYRLNVTNTVG YRFGGEVYGYGSDDEFKRA-----FGENSPYKK 552
D 553 -----AGEDHCNYOGSSSNYRDCVKRLIKGNKYFAARNMALGK 592
QY 553 HCNQSCGIYEPVLKYGKRRANHSSVISI-----SADFGDYTFMPF-----ASY--SRTHRM 600
```

Db 593 YVDLGLGIRVDV-----SRKANESTISVGKFKNFSWNTGIVIKPTWELDLISYRLSTGFRN 648
QY 601 PNQIEMFYFOIG---DSGVHTALKPERANTWQFCFNTRYKGLKLDQDDTLGLKLVGYSRRI 657
Db 649 PSFAEMYGWRYGGKNDVEYVYKFKPETSQNQEF-----GLALKGDFGNIET-----SHF 697
QY 658 DNIYHNHYGKWDNLNGIPSWSGTGLAYTIQHRNFKDKVHKHGFLELALNDYDGRFTNL 717
Db 698 SNAYRNLIAPAEELSKN-----GTGKNGNYGYHNAQNAKL--VGVNITAOQLDFNLWKRI 749
QY 718 SYAYOKSTQPTNFSASESPNNASKEDOLKOGYGLSRVSAL-----PRDY-----GRLEV 767
Db 750 PYGWY-----ATFANRVKVKDQ-KINAGLASVSYLFDATQPSYIIIGLVVDHI 798
QY 768 GTRWLGKMLTGLGAMRYFGKSIKIRATAPERYIDGTNGTNSNRQLGKRSIKQTETLARQP 827
Db 799 SNTW-----GVNATFTQS-KAKSQNELLGKRALGNNS-----RDVSKSTRKULTRA 842
QY 828 LIFDFYAAYEPKKNLFRABVKNLFRIDRYIDPLDAGNDAATQRYSSFDPKDKDEEVTGN 887
Db 843 HILDVSGYMYMANKNIMRLGIYLNFRYV-----TWEAVRQTAQGANV 886
QY 888 ADKTLCKNGYGGTSKSVLTNF-ARGRTFLITMSYKF 922
Db 887 QHQN- -GNY-----TRYASGRNYTLTKMF 912

RESULT 9
HGBA_HAEIN STANDARD; PRT; 1013 AA.
AC Q9KIV2:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin binding protein A precursor.
GN HGBA,
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=NTHI N182;
RX MEDLINE=20316037; PubMed=10958226;
RA Cope L.D., Hrkai Z., Hansen E.J.;
RT "Detection of phase variation in expression of proteins involved in
RT hemoglobin and hemoglobin-haptoglobin binding by nontypeable
RT Haemophilus influenzae."
RL Infect. Immun. 68:4092-4101(2000).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND
CC IS REQUIRED FOR HEME UPTAKE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAGING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF221059; AAF80176.1; -
CC InterPro: IPR000531; TonB_boxC.

Pfam: PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor; Repeat.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1013 HEMOGLOBIN BINDING PROTEIN A.
FT DOMAIN 26 57 8 X 4 AA TANDEM REPEATS OF Q-P-T-N.
FT REPEAT 26 29 1.
FT REPEAT 30 33 2.
FT REPEAT 34 37 3.
FT REPEAT 38 41 4.
FT REPEAT 42 45 5.
FT REPEAT 46 49 6.
FT REPEAT 50 53 7.
FT REPEAT 54 57 8.
FT SITE 67 74 TONB BOX.
FT SITE 996 1013 TONB C-TERMINAL BOX.
FT SEQUENCE 1013 AA; 116260 MW; 769964335A4ED3C1 CRC64;
Query Match 4.9%; Score 241.5; DB 1; Length 1013;
Best Local Similarity 19.2%; Pred. No. 1.2e-07;
Matches 210; Conservative 145; Mismatches 385; Indels 351; Gaps 49;
QY 40 LEDVHVK-----AKRVPKDKKVFVDARAVSTRQDIFKSSSENLDNIVRSIPGAFTQODK 92
Db 66 LEQINVSSTENSDSKTPPKIAETVKTAKTLERQ-----ANNIKDIVKYETGV----- 114
QY 93 SSGIVSLNIRGSGF-----GRVNTMVDGITQTFYSTSTDAAG-----GSSQF 137
Db 115 --TVVEAGREGQSGFAIRGVNENRVAINDGLRQA--ETLSSQGFELFCYGNFNTRN 170
QY 138 GASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVVGQNNNTYGLLKLGLTGN 197
Db 171 GAEIET--LKEVNITKANGSIKSGSLGSGSVIYKTDARDYLLNKDYVYSYKKGYATEN 228
QY 198 STGNMAAAGARKWLESASGVLYGH-----SRRSVAQNYRVGGG-- 240
Db 229 NQSFNTLTLAGRYKKFDVLYVVTTSRNGHELENYGKYKNYNDKIQGRREKADPYKIQDST 288
QY 241 -----QHIGNFGAYLERRKQ-----RYFVQEGGL-----KFSNSN 270
Db 289 LLKLSFNPTENHRETLAADLYEHRSGODLSYLYLYLTPDLPEVDSRHTNDKTRHNI 348
QY 271 SGKWERDFQPYWKTKWY-----QKYNPQELQKIEGHDKSWRENLAPOQDITPIDPSSL 326
Db 349 SFSYENFSQTPFWDTLKITEFSKQIKTRATDEYCD-----GVRYCEGTANPAGL 399
QY 327 KQOSAGNLFKLEVDGVFNKYTAQFDRLN--TKIGSRKIINRYOFNYGLSLNS----- 377
Db 400 KLN-GEITR--RDGT-----PLOFKEINNTTPNSNSNKOKTYDFSKLIDTNGEIESGI 452
QY 378 -----YANLNTAAYNSGRQKYPKGS--KFTG----- 402
Db 453 TRSNDTWDYCSIFDCENPGKMKVAECKTYRYDGTWKNNVQLEKVLNGKEPARINNGT 512
QY 403 -----W-----GLLKDFETY-----NNAKILDNNATTFR 427
Db 513 RGKTFPLPSSLYLERLWQEROLDTNTQQLNLDLTDFKTRVHEHNLQIGSSYNTMKR 572
QY 428 LPRE-----TELQ-----TTLG--FNYFHN-----YGNRF-----PEELGLFFD 461
Db 573 MVARAGYDATDVOMWAKRTLTGTRFDLKNEEIVETCATTFGNNAFLCPRVDP----- 625
QY 462 GPQDNGLYSLGRFGKDGKLLPOKSTIVQAGSOYFNTFVDAALKKIDYRLNNTV 521
Db 626 -----FSY-----LLPIK-----TKESVYLFDNVITD-----YLSFDL 655
QY 522 GYRFGG--EYGTGYG-----SDDEPKRAF-----GENSPYKKHCNOSCYIEVPLK 566
Db 656 GYRYDNHYQPKYKHGVTPKLPDDIVKELFPLKSGQNNDAAEVKKNVQENIDYIAQNK 715


```
Qy 567 KYGKKRANHNSVSISADGDFYFMPFASYSRTHRMPNQEMYS-QIGDSGV--HTALKPE 623
Dy 716 KNTWQFGNTYKGLLKQDDTLGLKLVGYSRSDINHYHNVYKWKWDLNIGNPSWSTG 683
Qy 724 RYATWQFGNTYKGLLKQDDTLGLKLVGYSRSDINHYHNVYKWKWDLNIGNPSWSTG 683
Dy 773 IAKTKEAFT-----LHDDWGFISTLFTKNTYRDFIDLVIKGEREFVGNPN--NRGK 824
Qy 684 LAY-TIOHRNFKDKVHKHGLELNYDGR-----FFNLVYAYOKSQTPNFSASE 735
Dy 825 ISFTTFQIN-RDSAVVKGIEINSKVFLGMAKFMGDFNLVYKTYQKGRWDGNI----- 878
Qy 736 SPNNASREDOLKQGYLSRVSALPRDYGRLEVGTWNLGNKLTILGGMARYFGKSTRATAEE 795
Dy 879 -PMNAIOPKTMVYGLGYDHP-----QKGFENFYTT--- 908
Qy 796 RYIDGTNGGNTSVNRQLGKRSIKQET-----LARQPLIDFYAAYEPKKNLIFRAEVKNL 851
Dy 909 -HVASKNPEDTYDIYAKDK---NOTNTSIKWSKSYTILDLIGYVQPIKNTIRAGVYNL 964
Qy 852 FDRYIDPLDAGNDAATORYSSDPKDKDEEYTCNADKTLNCKYGGTSKSVLTNEARG 911
Dy 965 TNKXII-----TWSARSIRSGFSNVIDQK-----TGGINRNFYAPG 1002
Qy 912 RTFLITMSYKF 922
Dy 1003 RNYKMSVQFEF 1013

RESULT 10
HGBC_HAEIN STANDARD; PRT; 993 AA.
AC Q9KIV0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin binding protein C precursor.
GN HGBC.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBL_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTHI N182;
RX MEDLINE=20316037; PubMed=10858226;
RA Cope L.D., Hekal Z., Hansen E.J.;
RT "Detection of phase variation in expression of proteins involved in
hemoglobin and hemoglobin-haptoglobin binding by nontypeable
Haemophilus influenzae.";
RL Infect. Immun. 68:4092-4101(2000).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND
IS REQUIRED FOR HEME UPTAKE.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAINING.
ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
Dy EMBL; AF221060; AAF80178.1; -;
```

```
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor; Repeat.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 993 HEMOGLOBIN BINDING PROTEIN C.
FT DOMAIN 26 37 3 X 4 AA TANDEM REPEATS OF Q-P-T-N.
FT REPEAT 26 29 1.
FT REPEAT 30 33 2.
FT REPEAT 34 37 3.
FT SITE 47 54 TONB BOX.
FT SITE 976 993 TONB C-TERMINAL BOX.
FT SEQUENCE 993 AA; 113616 MW; A551BF3B2C641612 CRC64;

Query Match 4.9%; Score 241; DB 1; Length 993;
Best Local Similarity 19.8%; Pred. No. 1.2e-07;
Matches 222; Conservative 156; Mismatches 414; Indels 328; Gaps 51;

Qy 3 SSRLRPICILMGVMLYHHSYAEDAGRAGSEAQI-----QVLEDVHVK-----AK 48
Dy 2 ANFLANLAYSVMLGLTAGVAYAAQPTNPTNQNNSVSEQLQINYSGSTENSDTK 61
Qy 49 RVPKDKVFTDARAVSTRODIFKSSSENLDNIVRSIPCAFTQODKSSGIVSLNIRGDSGF- 107
Dy 62 TPPKIAETVKTAKTLEREQ-----ANNKIDIVKETGV-----TJVAGRFQSGGPA 108
Qy 108 -----GRVNTMVDGITQTFYSTDAGRAG-----GSSQFGASVDSNFIAGLDVVK 153
Dy 109 IRGVDENRAVINIDGLRQA--ETLSSQGFKELEFEGYFNFNTRNGAEIET--LKEVNITK 164
Qy 154 GSFSGSAGINSLAGSANLRTLGLVDVVQGNNTYGLLLKGLTGNTSTKGNMAAIGARKWL 213
Dy 165 GANSIKSGSGSLGSGVIYKTKDARDYLLNKDYVYSYKGYATENNOSFNTLTLAGRYKKF 224
Qy 214 ESGASGVLYGHSRRSVAQNYRVGGGGQHIGNF---GAEYLERKKQR-----YFVQGG- 264
Dy 225 DA-----LVVTTNRN-----GHELENDYKNADSLTQCKKREKADPYKIEQDST 268
Qy 265 -LKFNSNGKWER-----DF-----QRPYWKTKWYQKYNDOELQYKYTEGHDKSWREN 311
Dy 269 LLKLSFNPTENHRTLAADLYEHRSGQDLSYTLKYQSRGNETPEVESR--HTNDKTKRRN 327
Qy 312 LAPQYDITPIDPSLSKQOSAGNLFKLEYD-----GVF----- 343
Dy 328 ISFSYE-----NFSQTPFWDTLKLYSDQRIKTRARTDEYCDAGVTHCQGTENPTGLK 380
Qy 344 -----NKYTAQFRDLNLTIGSRK-----IINRNYQ-----FNYGLSLSYANLNL 383
Dy 381 LTNGKITRRNGTDLQFKEKCTAAGTKYDFDTFIDTNNGEIVGKLNGLYANDTWYDCSL 440
Qy 384 TAYNSGRQYKPKGSKFTGW-----GLLKDFETYN-----AKIL 418
Dy 441 FDCDKGTCKMKVFEGNATYIGKWRSELETEVLNCKRFARIKDRKDKNNRNRKIKSIL 500
Qy 419 DLNATATERLPRTELQGT-----TLGFNYF---HN-EYKKN-----RPEELGLF 459
Dy 501 PSFPGYLERLWQRDLDTNTQQLNLDLTDFKTRVVEHNLYQSSYNTTKRNVNRAG-- 558
Qy 460 FDGPDQDNGLYSLGR-----FKG-----DKGLLPQ---KSTIVOPAGSQY 497
Dy 559 YDATDVQWAKRTLKSHNFFTGQDIVETCATTTSSSAWNAFLCPRVDPPEFSLPIKTKE 618
Qy 498 FNTFYFDAALKDIYRLNLYSTNTVGRFGG-EYTYGYG-----SDDEKRAF----G 544
Dy 619 KSYLFDNVVITD-----YLSFDLGYRYDNIHQYQIKHGVTPKLPDDIYKGLFIPLPK 673
Qy 545 ENSPTVKKHCNCSGIVPEVPLKYYKKRANNSVYSTADEFGYFMPFASYSRTHRMPNIQ 604
Dy 674 PNSNQVENVOQNLIDYIAKONKYY---KAHSYFASTIDTFSFLRLQLKYSKGFRAPTS 730
```


QY 605 EMYES-OIGDSGV--HTALKPERANTWQGFNTYKGLLKQDD-----TLGLKLVGYSRI 657
 Db 731 EMYETKHPDFTLLPNTNLKLEPAKTEAKT-----LHNDWDMFISLFXFTNYRDFI 784
 QY 658 DNYIHVYKQWDLNGLNIPESWSTGLATYIOHRNF-----KDKVHKHGFELNLYDYGR- 712
 Db 785 D-LVH-----KGRFKVKCPSEGEIEFTDTQINRDSAVVKGIEINSKVPFLGKM 834
 QY 713 -----FFNLSVAYOKSQPTNFSNPNASKEDQLKQGYGLRSALPRDYGRLE 766
 Db 835 AKPMGDGNLSYKTYOKGRMDGNI-----PMNAIQPKTMVYGLGYDHPES----- 878
 QY 767 VGRWLGNKLTGLGAMRYEKGSRATAERYIDGTNGGNTSNVROLGKRSIKOTEN--- 822
 Db 879 -----OKGFNFYTT-----HVASKPEDFYDIYAKDK-----NOTDTSIKW 915
 QY 823 LARQPLIFDYAAYEPKKNLIFRAEYKKNLDRRIYDPLDAGNDAATQRYYSFDPKDKDE 882
 Db 916 RSKSYTILDILGVQPKIKTLIRAGVYLNTRKYI-----TWDSARSIRSGTNSVIDQ-- 969
 QY 883 EYTCNADKTLNCKYGGTGSVLTNFARGRTFLITWSYKF 922
 Db 970 -----STGGINRFYAPGRNRYKMSVQFEF 993

RESULT 11
 HGP2_HAEIN STANDARD; PRT; 999 AA.
 AC P44809;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 2 precursor.
 DE precursor.
 GN HI0661.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 RN [2]
 RP REVISIONS.
 RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.,
 RA Peterson J., Hickey E., Dodson R., Gwinn M.;
 RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RX MEDLINE=20137488; PubMed=10675023;
 RA Langer H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
 RA Gray C., Fountoulakis M.;
 RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
 RL Electrophoresis 21:411-429(2000).
 RN [4]
 RP CONCEPTUAL TRANSLATION.
 RA Coudert E.;
 RL Unpublished observations (AUG-2001).
 CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
 CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED

CC FOR HEME UPTAKE (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
 CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
 CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
 CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
 CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
 CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO
 CC AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
 CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
 CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; TWO FRAMESHIFTS WERE
 CC INTRODUCED IN THE REPEATS REGION AND IN THE LEADER PEPTIDE TO
 CC MAXIMIZE THE SIMILARITY WITH OTHER ORTHOLOGS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32749; AAC22319.1; ALT_SEQ.
 CC TIGR; HI0661;
 CC InterPro; IPR000531; TonB_boxC.
 CC Pfam; PF00593; TonB_boxC; 1.
 CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 CC PROSITE; PS01156; TONB_BOX; 1.
 CC Outer membrane; Transport; TonB box; Multigene family; Signal;
 CC Receptor; Repeat; Complete proteome.
 CC SIGNAL 1 24 POTENTIAL.
 CC CHAIN 25 999 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-
 CC HAPTOGLOBIN BINDING PROTEIN 2.
 CC DOMAIN 26 53 7 X 4 AA TANDEM REPEATS OF Q-P-T-N.
 CC REPEAT 26 29 1.
 CC REPEAT 30 33 2.
 CC REPEAT 34 37 3.
 CC REPEAT 38 41 4.
 CC REPEAT 42 45 5.
 CC REPEAT 46 49 6.
 CC REPEAT 50 53 7.
 CC SITE 63 70 TONB BOX.
 CC SITE 982 999 TONB C-TERMINAL BOX.
 CC SEQUENCE 999 AA; 114690 MW; 1A17AAB220092B7D CRC64;
 SQ

Query Match 4.7%; Score 232.5; DB 1; Length 999;
 Best Local Similarity 19.8%; Pred. No. 4.2e-07;
 Matches 209; Conservative 144; Mismatches 415; Indels 285; Gaps 47;
 QY 40 LEDVHVK-----AKRVPKDKVFTDARAVSTRQDIFKSSNLDNIVRSIPGAFTOODK 92
 Db 62 LEQINVSGSTENSOTKTPPKIAETVKTAKTLEREQ-----ANNIKDIVKYETGV----- 110
 QY 93 SSGIVSLNIRGDSGF-----GRVNTWVDGITQTFTSTDTAGRAG-----GSSQF 137
 Db 111 --TWVEAGRFQSGFAIRGVDENRVAINIDLRAQ--ETLSQSGFKELFEGYGNFNTRN 166
 QY 138 GASVDSNFIAGLDVVKVGSFSGSAGINSLAGSNLRTLGVDVVGQNNNTYGLLLKGLGTN 197
 Db 167 GAETET--LKEVNITGADSIKNGSLGGSVIYKTKDARDYLINKDYIYYSYKGYATEN 224
 QY 198 STKGNMAAIGARKWLESASGVLYGH-----SRSVAQNYRVGGGG- 240
 Db 225 NQSFDTLTLAGRYKKFDVLVVTTSRNGHELENYKYNKDIQKKREKADPYKIEODST 284
 QY 241 -----QHIGNFGAEYLERKKQ-----RYFVQEGG-----LKFN 268
 Db 285 LLKLSFNPTENHRFTFAADLYEHRSGQDLSYTLKY--QRSGNETPEVDSRHTNDKTKRR 342
 QY 269 SNGSKWERDFQRYWTKYQKYNDDQELQ-----KYIEGHDKSWRENLAPO 315

Db 343 NISFSYENFSPFPWDT-LKLTYSO-QRIKTRARDEYCDAGVRHCEGTDNPTGLKVT-N 399
QY 316 YDITPIDSSLKQOAGNLFKLE---YDGFENKNTAQFRDLNLTGKSGKIINRNQYENG 372
Db 400 GKTRRDSGLQFEKNNKNTAKSSDKTYD--FKK-----FIDYDKRVDDKLVNN----- 447
QY 373 LSLNAYANLTAAYNSGRQYKPGSKETGNG-----LLKDPETYNNAKILDLNNTA 424
Db 448 -PSDTWYDCSIFENCENNAKIKVFKNYNYGYDGKWKVEYDLEIKELNGKKFAKIKDNRKI 506
QY 425 TFRLLP-----RETELDT-----TLGFNYPHNE-----YG-----KNRPFEE 455
Db 507 KSILPSSGYLERLWQERDLDTNTQQLMDLTKDFKIHIEHNLQYGSYNTAMKVMNR 566
QY 456 LGLFFDGPD-----QDNGLSYLGR-----FKGDKGLLPQ-----KSTIVQAPAGSOFVN 499
Db 567 AG--NDASDVQWATPTLGEDSWTGKPHTCATTYEWANLCPRPVDPPEFSYLLPIKTCKS 624
QY 500 TFYDAALKDIYRLNYSNTVGRFGG-EYTGYYG-----SDDEPKRAF-----GEN 546
Db 625 VYLFDFNFIYD-----YLSFDLGYRYDNIHYOPRYKHGHTPKLPDDIVKGLFIPLPNSN 679
QY 547 SPTYKHKCNQSCGIVPEVLKRYGKRRANHSVISADEGDFYFMPFASVSRTHRPNQEM 606
Db 680 SDPNKVENVOQNI-DYIAKONKYYKAHSYFVSTIDPTSELRQLKYSKGFRTPTSDM 738
QY 607 YFS-OIGDSGV--HTALKPERANTWQGFNTYKKGKLLQDDTLGKLVGYRSRIDNYIH 663
Db 739 YFTFKHPDFTLPLNTDLKPEIAKTAFT-----LHNDWGFISTSLFKTNKYNFIDL 792
QY 664 VYGKWDLNGNIPSWVSTG---LATIQHNEFKDKVHKHGFLELNYDGR----- 712
Db 793 IFKKQETFK-----VGSGRGETPLPSLYQNIINRDNASLKGIEINSKVLGKMAKPMFG 846
QY 713 FFTNLSYAYOKSTOPTNFSDASESPNASKEDQLQGYGLSRVSALPRDYGRLEVGTRWL 772
Db 847 FNLSYKVTYQKGRMNGNI-----PMNAIQPRTVYGLGYDHPN----- 884
QY 773 GNKLTGLGAMRYFKSIRATAERYIDTNGGNTSNVRQLGKRSIKOTETL---AROPLI 829
Db 885 -----HKGFDFYTT-----HVASKNPEDTYNM--FYKEENKDKSTIKWRSKSYTI 928
QY 830 FDFVAAVEPKNLFRAEVNLFDRYIDPLDAGNDAATORYYSSFPDKDKDEVTGNAD 889
Db 929 LDLGYVQPKNLIRAGVNLNRKYI-----TWDSARSIRSFCTSNVDO----- 975
QY 890 KTLGNGYKGTSKSVLNTFARGFRFLITMSYKF 922
Db 976 -----STGLGINRFYAPGRNKKMSVQFEF 999
RESULT 12
HGPI_HAEIN STANDARD; PRT; 1063 AA.
AC P44795;
DT '01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 1 precursor.
GN HI0635.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=2013748; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
RN [3]
RP CONCEPTUAL TRANSLATION.
RA Couderc E.;
RL Unpublished observations (AUG-2001).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAGING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO
CC AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE SEQUENCE WAS
CC ELONGATED IN THE N-TERMINAL SECTION, THEN A FRAMESHIFT WAS
CC INTRODUCED IN THE REPEATS REGION TO MAXIMIZE THE SIMILARITY WITH
CC OTHER ORTHOLOGS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32746; AAC22294.1; ALT_SEQ.
DR TIGR: HI0635;
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 2.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor; Repeat; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1063 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-
FT HAPTOGLOBIN BINDING PROTEIN 1.
FT REPEAT 26 53 7 X 4 AA TANDEM REPEATS OF Q-P-T-N.
FT REPEAT 26 29 1.
FT REPEAT 30 33 2.
FT REPEAT 34 37 3.
FT REPEAT 38 41 4.
FT REPEAT 42 45 5.
FT REPEAT 46 49 6.
FT REPEAT 50 53 7.
FT SITE 63 70 TONB_BOX.
FT SITE 1046 1063 TONB_C-TERMINAL_BOX.
SQ SEQUENCE 1063 AA; 121160 MW; 370CB515523F2788 CRC64;

Query Match 4.7%; Score 229; DB 1; Length 1063;
Best Local Similarity 19.0%; Pred. No. 7.6e-07;
Matches 211; Conservative 152; Mismatches 402; Indels 344; Gaps 47;

```
QY 41 EDVHVAKRVPKDKVFTDARAVSTRODIEKSSSENLDNIYRSIPGFTQODKSSGIVSLN 100
Db 72 ENINVKEKKVGETQ-----ISAKKIAQO-----ADSRD-LVRYETGTVVETGRTGASGYA 123
QY 101 IRGDSGGRVNTVDGITQTFYSTSDAG-----RAGSSQFCASVDSNFIAGLDVKVKG 154
Db 124 VRGDE-NRVGIMVGLRQA--ETLSSQGFKEFEGVGNFNTRNSIEIENVKTATITKG 180
QY 155 SFGSAGINSLAGSNRLTGVDDVVOGNTYGLLLKGLTGTNSTKGNAMAAIGARKWLE 214
Db 181 ADSLKSGLGSGVETPDKARDYLDKDYLSYKRGYQYTMNQNLTTLTLAGRSKFFD 240
QY 215 SGASGVGLYCHS-----RRSVAQNYRVGGGQGHGNEGAELYER 253
Db 241 ILIIDTRDGHENIYDKIYPNKAQDLRAVGPTRKADPYQITRQSTLI-KLGFQPNEN 299
QY 254 RKQRYFQOEGGLKFNSSNGKWERDFQRYWKTKWYQYNDPOELQKYLEGHDKSWRENLA 313
Db 300 HRLSVALDDSTLETKGIDLSYA--LRPY-STANNEXYG-----ERII--NDQSKRKNIQ 348
QY 314 POYDITPIDSSLSKQOQAGNLFKLEYGVFNKNYTAQFRDLNTKIGSRKIIN--RNVQFNY 371
Db 349 FSYE-----NFSQTPFWDHILKLSY-----SSOKITNKARSDEYCH 383
QY 372 GLSLNSVAN--LNLTAAYNSGRQKYPKG---SKFTGWGLLKDFETYNNAKILTD----- 419
Db 384 QSTCNGVSNPQGLHLVEKGVYKIKDYGGELESKEIGWS-----HEFKNSKGEDADKDI 438
QY 420 -----LNNATFRLPRETELTQTTLGFNYFHE-----YKG----- 449
Db 439 SQRSSLDVLINCEKDCSKKFRLYQYEDNSSEKYYDDRELEVGLPLNGKKGKPLK 498
QY 450 -----NRPEELGLFDPG-----DOD----- 466
Db 499 KGKTPSNWGPQETARELFPKSYGYSTDFVNDRLNTHTOQIKLDDKEFLHWTQHLK 558
QY 467 -NGLYS-----YLGFRKDGKLLLPQKST---IVOPAGSQYF 498
Db 559 YGGLYEKTLKSMVNHQYNTAANVOWWADYFFCARAKG--GNLGEKTPHPNVSVAGCVNG 616
QY 499 NTFVFDAAALKKDIYRLNSTNT-----VGVRF----- 525
Db 617 TPLHSD--IGKDYLIPIVTKNNVLYFGDNVOLTSLWGLDLNLYRDNHVKYLPGYDEKTPV 674
QY 526 -GGEYTYGSDDEKFRAGFENSTYKHKHNSQSGIYEP-----VLKKYKKRRANNH 576
Db 675 PGGIAGIFVPFNEKQVYVGYVPSGYKDCRYNTECYKKNFEENLALLLRKTDYKH-HSY 733
QY 577 SVTSADPGDYFMPFASYSRTHMPNIOEMY-----FSQIGDSGVHTALKPERANTWQ 629
Db 734 NLGLNLDPTDLWRVOLKRYANAFRAPTDEIYMTFKHPDFS-----IGPNTNLKAETAKTE 789
QY 630 FGFTYRKGLLKQDDTLGLKLVGYRSRIDNYIHNVYKQWMDLNGNIPSWVSSTGLAYTIQ 689
Db 790 VAFTFYKENSY-----LTLSAFQSDYRNFIDLVEFK-----NKQID--KGSIAIEPFY 835
QY 690 HRNFKPKVHKHGFLELNYOYGRFTNL-----STAYOKSTQPTNFSDAESPNNASK 742
Db 836 QNQRNDQARVGBIEASRLMGDLFEKLGQFHLGYKLTQK-----GRIK 880
QY 743 EDQLKQGYG-----LSRVSAIPRDYGRLEVTGRNLGNKLTIG--CAMRYFGKSTRAT 792
Db 881 DNKLRSYAEFLKLNPOYTATIASODQPMNALQPTT-----SVYINIGYDAPSKKMGMDYIT 936
QY 793 --ABERYIDGTNGCNTSNVR-----OLGKRSIKOTETLAR--OPLIFDFY 833
Db 937 DVAAKKADKSFNSQWTSWVRKENIYERTVPAQANGKVDKDRGLWNRNRYTVIDTI 996
QY 834 AAYEPKKNLIFRAEVKNLFDORRYIDPLDAGNDAATQRYYSFDPKDKDEEYVTCNADKTLK 893
Db 997 AYWKPIKNLFTAGVYNLTKKYL---TWDSARSVRHLG-----TINRVET-- 1039
QY 894 NGKYGGSKSVLTFNFRGRFTLITMSYKF 922
```

```
Db 1040 -----ATGKGLNRFYAPGRNRYMSVQFEF 1063
RESULT 13
LBPA_NEIMA
ID LBPA_NEIMA STANDARD; PRT; 944 AA.
AC Q9JTK4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Lactoferrin binding protein A precursor.
GN LBPA OR NMA1739.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL162757; CAB84967.1; -.
CC InterPro: IPR000531; TonB_boxC.
CC Pfam: PF00593; TonB_boxC; 1.
CC PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
CC SIGNAL 1 27 POTENTIAL.
CC CHAIN 28 944 LACTOFERRIN BINDING PROTEIN A.
CC SITE 827 944 TONB C-TERMINAL BOX.
CC SEQUENCE 944 AA; 105732 MW; CE06B6192E74AE3E CRC64;
Query Match 4.6%; Score 226.5; DB 1; Length 944;
Best Local Similarity 19.5%; Pred. No. 9.4e-07;
Matches 210; Conservative 155; Mismatches 379; Indels 335; Gaps 52;
QY 23 SYADAGRAGSE-AQIQVLEDVHVAKRVPKDKKVFVDARAVSTRQDIFKSSENLD---- 77
Db 22 SYANPETAADAAQOSLSKEITVRAAKVGRSKEATGL-----GKIVTSETLNKEQV 75
QY 78 ----NIVRSIPGAFQODKSSGIVSLNIRGDSGFRVNTWVDGITQ-----TFYSTSDAG 129
Db 76 LGIRDLTRYDPGVAVVEQNGCAGSGYSIRGVDK-NRVAVSVDGVGAQIQAFVQGSLSYCG 134
QY 130 RAGSGSFGASVDSNFTAGLDVVKGSFGSAGINSLAGSANLRLGLVDVDDVVOGNTTGLL 189
Db 135 GRGSGGAIN-EIEYENISTVEIDKAGSSDHGSGALGAGAVAFRTKEADLSDGKSWGIQ 193
QY 190 LKGLTGTNSTKGNAMAAIGARKWLESASGVLYGHSRVAQNYRVGGGQGHGNEGAEL 249
Db 194 AK--TAYGSKNRQPMKSLGAGFSKDGWEGLLIRTEROGRTRPHGDIADGVE----YIGD 247
```



```
QY 360 RKIINRYQENYGLSL-NSYANLNLTAAVNSGRQKPKGSKFTGWLLKDFETYNNAKIL 418
Db 307 QQIPDQSTVENYLYTPDNPYLNTHIALYNNKTIKEQR--KVSQ--VKD-QTKLTRGI 360
QY 419 DLNNTATFRLPRETELQTTLGFNFYHNEYKKNRFPPELGLFFDGDODNGLYSYLGFRKG 478
Db 361 NLNRSSEL-----SHISFYGVDMRDKIRTER-----G 389
QY 479 DKGLLPQKSTIVQAGSQYFNTFYDAALKDILRLNYSTNTVGYRFGGEYTGYSDDDE 538
Db 390 TNG-----SDAKFRADPY--NANSNTTGVYLIAHIP-----418
QY 539 FKRAFGEN---SPTYK-KHCNQCSCIYEPVLKYGKRRANNHVS-----ISADFGDYFM 589
Db 419 ---LFGEKLLVSPSVRYDHYDTS-----SKTVYKDNHLSPATKLTWIVTNWLD-- 465
QY 590 PEASYSRTHRMPIQEMYSFQIGDSGVHTA-----LKPERRANTWQFGF 632
Db 466 -TAKYNEAFRAFSQMERFV-----SGAHFGANTLGLDHIINRFVANPLRPETAKNKEITA 519
QY 633 NTKYKGLLKODDTLGLKLVGYRSRIDNYIH-NVYKQWMDLNGNIPSWVSSTG-----683
Db 520 NLHFDLSFKQGDKFKEATYFRNDVKDFINLKIF-----NDAKTSASAGANPNTGA 571
QY 684 -LAYTQHRNFKDKVHKHGFELNLDYDGR--FFTNLSYAYQKSTQPTNFSASESPNNA 740
Db 572 LLPKNSQYQNTN-ARLSGIELQAQYQTERLTLFTNY-----GS 609
QY 741 SKEDQLKQGYGLSRVSALPRDYGRLEVGTRW--LGNKLTGGAMRYFGKSIKIRATAERYI 798
Db 610 TKGKDKDSGALSNIAA-----SKIGVGVNVALVKDFVTGATVTHAAO-----654
QY 799 DGTNGGNTSNVRLGKRSIKOTELARQPLIFDFYAAYPEK-----KNLIPRAEVKNLFR 854
Db 655 -----RRVPKDHVSYPYSYILTDLATYAPLKGEWKNLRLDFALENLFR 699
QY 855 RYIDPLD-----AGNDAATQRYIS 873
Db 700 KYQPAFSLMEGTGRNAKISAVS 722
```

RESULT 15

```
HXC2_HAEIN
ID HXC2_HAEIN STANDARD; PRT; 725 AA.
AC P45357;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin utilization precursor.
GN HXUC.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DL42 / SEROTYPE B;
RX MEDLINE=95270579; PubMed=7751272;
RA Cope L.D., Yegor R., Mueller-Eberhard U., Hansen E.J.;
RT "A gene cluster involved in the utilization of both free heme and
heme/hemopexin by Haemophilus influenzae type b.";
RL J. Bacteriol. 177:2644-2653(1995).
CC -!- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
CONCENTRATIONS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -!- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE D OF
H. INFLUENZAE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U09840; AAA87059.1; -.
InterPro: IPR000531; TonB_boxC.
Pfam: PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; signal; Receptor.
FT SIGNAL 1 19 OR 21 (POTENTIAL)
FT CHAIN 20 725 HEME/HEMOPEXIN UTILIZATION PROTEIN C.
SQ SEQUENCE 725 AA; 80838 MW; FC7886E020CBCE CRC64;

Query Match 4.3%; Score 211.5; DB 1; Length 725;
Best Local Similarity 19.9%; Pred. No. 5.8e-06;
Matches 133; Conservative 130; Mismatches 318; Indels 289; Gaps 46;

```
QY 31 AGSEAQIQVLEDVHVAKRYP-----KDKKYFTDARAVSTRODIFKSSSENLDNI 79
Db 17 ANALAQSVELDSINVIAIRDPSPFAVTPKOSKDSLLSKQATSVAA-----ALEDIPNV 70
QY 80 -VRSIPCAFTQODKSSGIVSLNIRGDSGFRVNTWVDGIIQTIFYSTSDAGRAGGSSQFG 138
Db 71 DIRGGSRSIAQKP-----NIRGLSD-NRVQVIDGVQRNF-----DLAHRG--SYF- 113
QY 139 ASVDSNFIAGLDVYKGSFSGSAGINSLAGSANLRTLCVDDVVOGNNTYGLLLK-GLGTN 197
Db 114 --LPMSLIQIEVWIKGSSSLWSSGALGVVAMTNPALDLLKNNDFGVKIROGYOTAN 171
QY 198 STKGNAAMAATGARKWLESAGVLYGHRSRVAQNYRVGGGQGHIGNFGAELYERRKQR 257
Db 172 NLSEKRVSVFAANDKFD--VLISGFYNN-----ADNLTGKG-----NKLNTT 212
QY 258 YFVQEGGL-KFNSNSGKWE-RDFQR---PYWKTKWYQKYNPDQELQKYIEGHDKSWREN 312
Db 213 AYKQFGGLAKFG-----WQIDANRVLSHRETRFKOTAPSNNEV-----ENE 255
QY 313 APOYDIT-----PIDPSSLKQOASAGNLFKLEYDGVFNKYTAQFRDLNLTIGS-----359
Db 256 LTNEQITDQIREPHKPNNGSPPKAKPSQEEF---YSGV-----KTRFQSVSYLT 301
QY 360 -RKIINRYQENYGLSL-NSYANLNLTAAVNSGRQKPKGSKFTGWLLKDFETYNNAKI 417
Db 302 DQIQPDQSTVENYLYTPDNPYLNTHIALYNNKTIKEQR--KVSQ--VKD-QTKLTRG 355
QY 418 LDLNNTATFRLPRETELQTTLGFNFYHNEYKKNRFPPELGLFFDGDODNGLYSYLGREF 477
Db 356 INLRNSEL-----SHISFYGVDMRDKIRTER-----384
QY 478 GDKLLPQKSTIVQAGSQYFNTFYDAALKDILRLNYSTNTVGYRFGGEYTGYSDD 537
Db 385 -----GTNNKDAQFRADPY--NANSNTTGVYLIAHIP-----414
QY 538 EFKRAFGEN---SPTYK-KHCNQCSCIYEPVLKYGKRRANNHVS-----ISADFGDYF 588
Db 415 ---LFGEKLLVSPSVRYDHYDTS-----SKTVYKDNHLSPATKLTWIVTNWLD-- 461
QY 589 MPFASYSRTHRMPIQEMYSF-----OIGDSGVHTALKKPERANTWQFGNTYK 636
Db 462 --TAKYNEAFRAFSQMERFVSGSHFGTSILGRNEINFKVANPLRPETAKNKEITANLHF 519
QY 637 KGLLKQDDTLGLKLVGYRSRIDNYIH-NVYKQWMDLNGNIPSWVSSTG-----LA 685
Db 520 DSLFKQGDKFKEATYFRNDVKDFINLKIF--DAKTNNTNASAGAGAGANPANGALLP 576
QY 686 YTTQHRNFKDKVHKHGFELNLDYDGR--FFTNLSYAYQKSTQPTNFSASESPNNAKSE 743
Db 577 TKSQYQNTN-ARLSGIELQAQYQTERLTLFTNY-----GSTKG 614
QY 744 DQLKQGYGLSRVSALPRDYGRLEVGTRW--LGNKLTGGAMRYFGKSIKIRATAERYIDGT 801
```

```
Db 615 KKDSEALSNIAA-----SKIGVGVNVALVKDKFTYGVATVTHYAAQ----- 656
Qy 802 NGGNTSNVRQLGKRRIKOTETLARQPLIFDFYAAYPEK----KNLIFRAEVKNLFDRIYI 857
Db 657 -----RRVPKDHVTPSYILTDLRATYAPLKGWKNLRLDFALENLFDRIYQ 704
Qy 858 DPLD-----AGNDAATQRIYS 873
Db 705 PAFSLMEGTGRNAKISAVYS 724
```

Search completed: July 24, 2002, 09:00:10
Job time: 401 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:52:04 ; Search time 68.49 Seconds
(without alignments)
2328.826 Million cell updates/sec

Title: US-09-762-926-4
Perfect score: 4904
Sequence: 1 MRSSFLKPCIFYLMGVMLY.....SVLTNFARGRTFLTMSYKF 922

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	4840	98.7	922	16	Q9JTN4		Q9jtn4 neisseria m
2	4768.5	97.2	921	16	Q9JR50		Q9jr50 neisseria m
3	1085	22.1	755	16	Q9PIV1		Q9piv1 campylobact
4	903	18.4	753	2	Q93052		Q93q52 shigella fl
5	902	18.4	753	2	Q93384		Q933s4 escherichia
6	459	9.4	885	16	Q92NX4		Q92nx4 rhizobium m
7	425	8.7	883	2	Q9X5P4		Q9x5p4 pseudomonas
8	425	8.7	891	16	Q9HYJ7		Q9hyj7 pseudomonas
9	425	8.7	989	16	Q9HUR6		Q9hur6 pseudomonas
10	415	8.5	916	2	Q9RHT4		Q9rht4 pseudomonas
11	396.5	8.1	899	2	P72473		P72473 serratia ma
12	351	7.2	848	16	Q9CKJ4		Q9ckj4 pasteurella
13	339.5	6.9	747	16	Q98L70		Q98l70 rhizobium l
14	312.5	6.4	782	2	Q93SH4		Q93sh4 bradyrhizob
15	308	6.3	784	16	Q9KLM6		Q9klw6 vibrio chol
16	278.5	5.7	739	2	O87381		O87381 haemophilus

ALIGNMENTS

RESULT 1

Q9JTN4 PRELIMINARY; PRT; 922 AA.
AC Q9JTN4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE OUTER MEMBRANE SUBSTRATE BINDING PROTEIN.
GN NNA1700.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holroyd S., Jagels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skellton J., Whitehead S., Spratt B.G., Barrell B.G.;
RT Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
RL Nature 404:502-506(2000).
DR EMBL: AL162756; CAB84928.1; .
DR InterPro: IPR000531; TonB_boxC.
DR PROSITE: PS01136; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 922 AA; 104290 MW; B45E683985DE20B6 CRC64;

Query Match 98.7%; Score 4840; DB 16; Length 922;
Best Local Similarity 98.7%; Pred. No. 3.9e-260;
Matches 910; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
Qy 1 MRSSFLKPCIFYLMGVMLYHHSYAEDAGRAGSEAQIQVLEVDVHKAKRVPKDKKVFDTA 60
Db 1 MRSSFLRKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLEVDVHKAKRVPKDKKVFDTA 60

```
QY 61 RAVSTRQDIFKSSNLNIVRSIPGAFQTDQKSSGIVSLNIRGDSGFRVNTWVDGITQT 120
DB 61 RAVSTRQDIFKSSNLNIVRSIPGAFQTDQKSSGIVSLNIRGDSGFRVNTWVDGITQT 120
QY 121 FYSTSTADAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
DB 121 FYSTSTADAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASVGLYGHRSRVAQNYRVGGGG 240
DB 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASVGLYGHRSRVAQNYRVGGGG 240
QY 241 QHIGNFAEYLERRKQRYFVQEGGLKFNSSNGKWERDFQRPYKTKWYQKYNPDQELQKY 300
DB 241 QHIGNFAEYLERRKQRYFVQEGGLKFNSSNGKWERDFQRPYKTKWYQKYNPDQELQKY 300
QY 301 IEHDKSWRENLAPOYDITPIDPSSLKQOQAGNLFKLEYDGVNKNKYTAQFRDLNKTIGSR 360
DB 301 IEHDKSWRENLAPOYDITPIDPSSLKQOQAGNLFKLEYDGVNKNKYTAQFRDLNKTIGSR 360
QY 361 KIINRNYQFNGLSLNSYANLNTAAYNSGRQKYPKSGKETGWLKDFETYNNAKLIDL 420
DB 361 KIINRNYQFNGLSLNSYANLNTAAYNSGRQKYPKSGKETGWLKDFETYNNAKLIDL 420
QY 421 NNTATFRLPRETELQTLTGLGFNYFHNHSGKWERDFQRPYKTKWYQKYNPDQELQKY 480
DB 421 NNTATFRLPRETELQTLTGLGFNYFHNHSGKWERDFQRPYKTKWYQKYNPDQELQKY 480
QY 481 GLLPQKSTIVOPAGSQYFNTFYDAALKDIIYRLNSTNTVGYRFGGEYTGYSDDDEFK 540
DB 481 GLLPQKSTIVOPAGSQYFNTFYDAALKDIIYRLNSTNTVGYRFGGEYTGYSDDDEFK 540
QY 541 RARGENSPYKHKCNQSCGIEPVLLKYGKGRANNHNSVLSADFGDYFMPFASYSRTHRM 600
DB 541 RARGENSPYKHKCNQSCGIEPVLLKYGKGRANNHNSVLSADFGDYFMPFASYSRTHRM 600
QY 601 PNIOEMFYSGIDSGVHTALKPERANTWQFNTYKGLKQDDTLGLKLVGYRSRIDNY 660
DB 601 PNIOEMFYSGIDSGVHTALKPERANTWQFNTYKGLKQDDTLGLKLVGYRSRIDNY 660
QY 661 IHNYYGKWDNLGNIPSWWSSTGLAYTIOHRNFKDKVHKHGFLELNDYDGRFFTNLSYA 720
DB 661 IHNYYGKWDNLGNIPSWWSSTGLAYTIOHRNFKDKVHKHGFLELNDYDGRFFTNLSYA 720
QY 721 YQKSTQPTNFSADSESPNNASKEDQLKQYGLSRVSALPRDYGRLEVGTRWLNKLTLLG 780
DB 721 YQKSTQPTNFSADSESPNNASKEDQLKQYGLSRVSALPRDYGRLEVGTRWLNKLTLLG 780
QY 781 AMRYFGKSIRATAERYIDCTNGNTSNVRLGKRSIKQETTLARQPLIFDFVAAYEPKK 840
DB 781 AMRYFGKSIRATAERYIDCTNGNTSNVRLGKRSIKQETTLARQPLIFDFVAAYEPKK 840
QY 841 NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEVTGNADKTLNKGYYGT 900
DB 841 NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEVTGNADKTLNKGYYGT 900
QY 901 SKSVLTNFAGRFTFLITMSYKF 922
DB 901 SKSVLTNFAGRFTFLITMSYKF 922

RESULT 2
QJR50 ID QJR50 PRELIMINARY; PRT; 921 AA.
AC QJR50:
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE PUTATIVE TONB-DEPENDENT OUTER MEMBRANE RECEPTOR (TONB-DEPENDENT
DE RECEPTOR).
GN TDFH OR NMB1497.
OS Neisseria meningitidis, and
```

```
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487, 491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IR1074;
RA Turner P.C., Thomas C.E., Stojiljkovic I., Elkins C., Kizel G.,
RA Ala Alden D.A., Sparling F.P.;
RT Identification and Characterization of Putative TonB-dependent Outer
RT Membrane Proteins among Pathogenic Neisseriae.
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Feden J.F., Dodson R.J.,
RA Nelson W.C., Winn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA Smith J.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL: AF227418; AAF73907.1; -
DR EMBL: AE002499; AAF41853.1; -
DR TIGR: NMB1497; -
DR InterPro: IPR000531; TonB_boxC.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 921 AA; 104222 MW; 67985B75EDF8819D CRC64;

Query Match 97.2%; Score 4768.5; DB 16; Length 921;
Best Local Similarity 97.7%; Pred. No. 3.6e-256;
Matches 901; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

QY 1 MRSSFLRPKICFYLMGVMLYHHSVAEDAGRAGSEAQIOVLVDVHVKAKRVPKDKVFTDA 60
DB 1 MRSSFLRPKICFYLMGVMLYHHSVAEDAGRAGSEAQIOVLVDVHVKAKRVPKDKVFTDA 60
QY 61 RAVSTRQDIFKSSNLNIVRSIPGAFQTDQKSSGIVSLNIRGDSGFRVNTWVDGITQT 120
DB 61 RAVSTRQDIFKSSNLNIVRSIPGAFQTDQKSSGIVSLNIRGDSGFRVNTWVDGITQT 120
QY 121 FYSTSTADAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
DB 121 FYSTSTADAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASVGLYGHRSRVAQNYRVGGGG 240
DB 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASVGLYGHRSRVAQNYRVGGGG 240
QY 241 QHIGNFAEYLERRKQRYFVQEGGLKFNSSNGKWERDFQRPYKTKWYQKYNPDQELQKY 300
DB 241 QHIGNFAEYLERRKQRYFVQEGGLKFNSSNGKWERDFQRPYKTKWYQKYNPDQELQKY 300
QY 301 IEHDKSWRENLAPOYDITPIDPSSLKQOQAGNLFKLEYDGVNKNKYTAQFRDLNKTIGSR 360
DB 301 IEHDKSWRENLAPOYDITPIDPSSLKQOQAGNLFKLEYDGVNKNKYTAQFRDLNKTIGSR 360
QY 361 KIINRNYQFNGLSLNSYANLNTAAYNSGRQKYPKSGKETGWLKDFETYNNAKLIDL 420
DB 361 KIINRNYQFNGLSLNSYANLNTAAYNSGRQKYPKSGKETGWLKDFETYNNAKLIDL 420
QY 421 NNTATFRLPRETELQTLTGLGFNYFHNHSGKWERDFQRPYKTKWYQKYNPDQELQKY 480
DB 421 NNTATFRLPRETELQTLTGLGFNYFHNHSGKWERDFQRPYKTKWYQKYNPDQELQKY 480
QY 481 GLLPQKSTIVOPAGSQYFNTFYDAALKDIIYRLNSTNTVGYRFGGEYTGYSDDDEFK 540
DB 481 GLLPQKSTIVOPAGSQYFNTFYDAALKDIIYRLNSTNTVGYRFGGEYTGYSDDDEFK 540
```



```

Db 430 -----YKACDSRVICVPGSYDIDKEGG-----FNPSVQLSAQVTPWLQPIGY 475
QY 595 SRTHMPNIQEMTFSGQDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGKLGVYR 654
Db 476 SKMRAFNIEQMFNSGGASMPFLPKPERAETWQAGFNIDTRDLLVEODALRFKALAYR 535
QY 655 SRIDNYIHN-----VYKWKW-----DLNGNIPSWSSSTGLAYTIQHRNF 693
Db 536 SRQNYIYSYSLVCSGRKCSMAEVIGNDWEGISDYSNNMIYVNSA-----584
QY 694 KDVHKGHELELYDYGRETFNLAYQKSTOPTNFSDASESPNNAKEDQLKQGYGLS 753
Db 585 SDVIK-GFELEMDYDAGFAFGRLSFSQOQTDQPTSTASY-----FGAG 628
QY 754 RVSAIPRDYGRLEVGTRWLNKNTLGLGAMRYFKSIRATAERYIDGTNGGNTSNVRQLG 813
Db 629 DMTLEPKRYMTLGTGVFFNALLTGLPIIKYTKARLSPDFQDEHTGA-----678
QY 814 KRSTKQFETLAROPLIFDEVAAYEPKKNLIFRAEVKNLFDORRYIDPLDAGNDAATQRYYS 873
Db 679 --IIKQ--DLPQPTIIDLYGTVEYNRLTLKLSVONLMNRDYSEALKNL-----725
QY 874 SFDPKKDEEVTGNADKTLCLNGYKGTGSKSVLTNFARGRTFLITMSYKF 922
Db 726 -MPEGLDETHPANS-----ARGRTWIFGGDIRF 753

RESULT 6
Q92NX4 ID Q92NX4 PRELIMINARY; PRT; 885 AA.
AC Q92NX4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PUTATIVE IRON/HEME TRANSPORT PROTEIN.
GN SMC04205.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=1021.
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Ramsperger U., Surzycki R., Thebaud P., Vandenbol M.,
RA Vorholter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL Science 293:668-672(2001).
DR EMBL; AL591789; CAC46612.1; -.
KW Complete proteome.
SQ SEQUENCE 885 AA; 95389 MW; BEE36641060DFBF4 CRC64;

```

```

Query Match 9.4%; Score 459; DB 16; Length 885;
Best Local Similarity 22.9%; Pred. No. 2.1e-17;
Matches 224; Conservative 122; Mismatches 332; Indels 300; Gaps 42;

QY 25 AEDAGRAGSEAQIOVLDEHVHAKRVPKDK-----VFTDARAVS-TRQDIFKS- 72
Db 128 AGDAGPCEGEGATLVRIYVTKTGRNANSAGAGFGTDPWYEEPASVSVSRDAVQR 187
QY 73 -SENLDNIVRSIPCAFTQDQKSGGI-VSLNIRGDSGFRVNTWVDGITQTFYSTSTDAGR 130
Db 188 AARNANDVLDSVAGVTSNRSEAQNPGIINVRLQDNQRTVMTIDGARQDFQ-----R 240

```

```

QY 131 AGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVVGQNTYGLLL 190
Db 241 AGHGAORVVYDTAFALRSVEYKGAAGVAGAGSLGAVNFTVTTADITIPDRDRGVEL 300
QY 191 KGTGTNSTKGNMAAICARKWLSEASGVLYGHRSRVAQNYRVGGGG-----241
Db 301 NAETGTNAYFYNG-SLIGAARFSE--DFSVLGGISRRKVG-DYDFGONGKSPLLDLAVT 355
QY 242 -----HIGNFGAEYLERRKORYFVQEGGLKFNKSGKWERDQRPYWKYTKYQY 291
Db 356 TAVDDSLFSLRLETFTGL-----LKV-----EGSPSDDFT--FDLSWLR-- 392
QY 292 NDPOELQKYIEGHDKSWRENLAPOYDITIPDPSLKKQOSAGNLFKLEVDGVFNKYTAQFR 351
Db 393 NDSEAIQGGVFGD--LRD-----DPQNYLNTVSSSEFWDPD-----SELI 432
QY 352 DLNATKIGSRKIIN---RNYQFNGLSNLSYANLNLTAAYNSGRQKYPKSGKFTGWLKLD 408
Db 433 DLKGRUNYRVNDELRYT-----PRLPITYAMTS-----FGG-----466
QY 409 FETYNNAKILDLNNTATFRLPRETELQTLGFNY---FHNEYGNKRPPEELGLFFDGPDQ 465
Db 467 -----SLDNTSRF-----ETAL-GDLSLNYGGEAYSNGKTTTPPLV-----D 503
QY 466 DNGL---YSLGRFKGDKGLLPQKSTIVQAGSQYFNTFYFD-----AALKKD 510
Db 504 DQGFDEAYGYKG-----LNPVGRSRMTSAFLNATLEHDDWLEWAGAGLRD 548
QY 511 IYRLNYSNTNVG-----YRFGGEYTYGSDDEFKRAFGENSPYTKKHNCQSCGIYEPVLK 566
Db 549 RYRLKGFTEVGGKPRYIVVPGVCYFYDDGEC--AYDEDPVY-----GGGEAVLE 598
QY 567 KYGKKRANN---HSVISADFGDYMPFASYSRTHRPNTQOEMYFSQIGDSGVHTA---- 619
Db 599 RVDIDKSGGALLPSARIAVMPFEGIQPFVYAHYRPPSVMEAL-----TSGGHPGDAIA 653
QY 620 -----LKERANTWQFGNTYKGLLKQDDTLGLKLVGRSRIDNYI--HNYYGKWW 670
Db 654 TYIPNPKLPERGRTWELGINIARDGLFTAGDSLRLTVYVDFRTIQDYITLGNFYFATFD 713
QY 671 LNGNIPSWSSSTGLAYTIQHRNFKDKVHKGFLELELYDYGRETFNLISYAYQKSTQPTNF 730
Db 714 KN-----LFQHVNLDDGDTMNGVEIEASYDMGSAVGSAYTKLTDYADTY 759
QY 731 S-----DASESPNNAKEDQLKQGYKSVSALPRDYGRLEVGTRWLNKNTLGGAMRYFG 786
Db 760 SYSQPTASGTPLAASGNTVPV-----SVLFVPPENKFTLDAGIRLFEKRLVGGRTY-- 812
QY 787 KSRATAERYIDGTNGG--NTSNVRQLGKRISKQETLARQPLIFDFAAYEPKKNLIF 844
Db 813 -----VSDSKPTVQLAGLFTAGYK-----VFDIYGSYSFSDSAKL 849
QY 845 RAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEEVTNADKTLCLNGYKGTGSKSV 904
Db 850 RLAINNVTDOYAPALCA-----FYYP-----871
QY 905 LTNPARGRTFLITMSYKF 922
Db 872 ----APGRTATVSLNFKF 885

RESULT 7
Q9X5P4 ID Q9X5P4 PRELIMINARY; PRT; 883 AA.
AC Q9X5P4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HEME RECEPTOR HASR.
GN HASR.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.

```

```
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20121752; PubMed=10658665;
RA Ochsner U.A., Johnson Z., Vasil M.L.;
RT "Genetics and regulation of two distinct haem-uptake systems, phu and
RL has, in Pseudomonas aeruginosa.";
RL Microbiology 146:185-198(2000).
DR EMBL; AF127223; AAD31013.1; -
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor.
SQ SEQUENCE 883 AA; 96942 MW; A29B07A3A6C31DCF CRC64;

Query Match 8.7%; Score 425; DB 2; Length 883;
Best Local Similarity 22.7%; Pred. No. 1.6e-15;
Matches 215; Conservative 114; Mismatches 382; Indels 238; Gaps 37;

Qy 30 RAGSEAQIQVLEDVHVKAKRVPKDKKVFDTARAVST--RQDIFKS--SENLDNIVRSIPGA 86
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 RVSQDDLQVQSPSV--ISAAR--PDDWVYQTPHSVSVIGREQIERNPPRHAADMLEETPGV 173

Qy 87 FTQODKSSGIVSLNIRGDSGFRVNTWVDGITQTFTYSTDAGRAGSSQFGA--SVDSNF 145
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 YSSVSQDDPGLSVNIRGIQDYGYNMSVDGMQRNYQ-----QSGHQRNGLTVDPDL 226

Qy 146 IAGLDVVKGFSGSGAGINSLAGSANLRTLGVDDVVQGNNTYG---LLLKGLTG--TNSTK- 200
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 227 LSEVVIDKGASSAMGGAGVIGGIANFRTLEARDLVRPKQVGGVRVLTSLGGDANGTHF 286

Qy 201 -GNAMAAGARKWLESGASGVLYGHSSRRSVAQNYRVGGGQGHIGNFGAEYLERKORYE 259
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 IGSAAFAIGTEVW-----DMLVAASERHLG--DYDPGTKGS--IG-----ELRTGAWF 330

Qy 260 VOEGGLKFNNSGKWERDFQRPYKTKWYKYNPQELQKIEGHDKSWRENLAPOVDIT 319
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 NPEAGORVK-----HSPVAYSGV-----MRSRLAKLGV 361

Qy 320 PIDPSSLKQOAGNLFKLEYDGVFNKYTAQPRDLNTKIGSRKIINRNQYFNGLSL--NSY 378
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 PQDQ---RLQFSYLTQVSYDDA--NMLNTENQALWEKLGSSDVAQNAFIDYGVAPDNPL 417

Qy 379 ANLNLTRAY--NSGRQYKPKGSKFTGNGLLKDFETYNNAKILDLNLTATPLRPTELQ 436
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 VDFRAKLYYVDNRNQOTLQGITPGYSITVQTDYTG---AQAOQSTFALDLSLTRA 473

Qy 437 TLGFNYPHNEYGKNR-----PPEELGLFFDGDQDQNGLYSLGRFKGDKGLLPQ 485
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 NYGLEFFYDVKVRPDSOSPRASTSAVGPAAEGM---TPKGDRLGSLFARLDYD----- 524

Qy 486 KSTIVQAGSQYFNTFFDAALKKDIIYRL-----NSTNTGVYRFGGEYTYGYSDEDF 539
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 525 -----YDDLNLNAGLRVDRYLRGDTGCFNARTILGTTRTDMPLQVAVD--- 570

Qy 540 KRAFGENSPTYKHCNOCSGIYEPVLKKYKKRANNSVSIADFG--DYFMPFASYSRTH 598
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 571 -REEGRFSPTF-----GLSVKPGVDMLQLFATYKGV 601

Qy 599 RMPNIQEWYSQIGDSG-----VHTALKPERANTWQFGENTYKGLLKQDDTLGLKLVG 652
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 602 RPAVATESLTITGRPHGGGAENMYNPNFLSPERSKAWEGFNVLENLWFSDDRGLGVAY 661

Qy 653 YRSRIDNVIHNVYKWNLDNGNIPSWSSSTGLAVTIQHRNFKDKVHKHGFELNLYDGR 712
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 FDTRVDDFIEMGMQ-----PPGCMAGIGNSAVNNL--DSTRPRGEYQLDYDAGL 713

Qy 713 FFTNLVSAYOKS-----TQPTNFSASESPNNASKEDQLKQYGLSRV--- 755
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 714 AYGOLSYTHMIGNSDFCSKTAWLGGVTVTKVSGRRPPVIMDRPDEQANAATHCSAVLGS 773

Qy 756 -SALPRDYGRLGVGTRWLNKNLILGGAMRYFGKSIRATAEERYIDGTNGGNTSNVROLGK 814
```

```
Db 774 AEHPMDRG-----SLTLG--MRFEDRLDVGARARYSEGYSVAGGATVSOAGV 820
Qy 815 RSIQQTETLARQPLIFDFYAAEYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAAQRYSS 874
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 821 YPADWKEY-----TWYDLYGSYRVSDELTTLRLAMENVTDRAYLVPLG----- 862
Qy 875 FDPKDKDBEVTNADKTLCKNGYKGTSKSVLT--NFARGRTFLITMSYKF 922
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 863 -----DVLAFTLGRGRTLQGTLEYQF 883

RESULT 8
Q9HYJ7 PRELIMINARY; PRT; 891 AA.
AC Q9HYJ7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HEME ACQUISITION PROTEIN HASR.
GN HASR OR PA3408.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004762; AAG06796.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 891 AA; 97864 MW; 126D6F41E7D66979 CRC64;

Query Match 8.7%; Score 425; DB 16; Length 891;
Best Local Similarity 22.7%; Pred. No. 1.6e-15;
Matches 215; Conservative 114; Mismatches 382; Indels 238; Gaps 37;

Qy 30 RAGSEAQIQVLEDVHVKAKRVPKDKKVFDTARAVST--RQDIFKS--SENLDNIVRSIPGA 86
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 RVSQDDLQVQSPSV--ISAAR--PDDWVYQTPHSVSVIGREQIERNPPRHAADMLEETPGV 181

Qy 87 FTQODKSSGIVSLNIRGDSGFRVNTWVDGITQTFTYSTDAGRAGSSQFGA--SVDSNF 145
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 YSSVSQDDPGLSVNIRGIQDYGYNMSVDGMQRNYQ-----QSGHQRNGLTVDPDL 234

Qy 146 IAGLDVVKGFSGSGAGINSLAGSANLRTLGVDDVVQGNNTYG---LLLKGLTG--TNSTK- 200
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 LSEVVIDKGASSAMGGAGVIGGIANFRTLEARDLVRPKQVGGVRVLTSLGGDANGTHF 294

Qy 201 -GNAMAAGARKWLESGASGVLYGHSSRRSVAQNYRVGGGQGHIGNFGAEYLERKORYE 259
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 IGSAAFAIGTEVW-----DMLVAASERHLG--DYDPGTKGS--IG-----ELRTGAWF 338

Qy 260 VOEGGLKFNNSGKWERDFQRPYKTKWYKYNPQELQKIEGHDKSWRENLAPOVDIT 319
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 NPEAGORVK-----HSPVAYSGV-----MRSRLAKLGV 369

Qy 320 PIDPSSLKQOAGNLFKLEYDGVFNKYTAQPRDLNTKIGSRKIINRNQYFNGLSL--NSY 378
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 PQDQ---RLQFSYLTQVSYDDA--NMLNTENQALWEKLGSSDVAQNAFIDYGVAPDNPL 425
```

```
Qy 379 ANLNLTAA--NSGRQKYPKSGKGTGWLKDKDFETYNNAKILDLNNTATFRLPRETELQ 436
Db 426 VDFKAKLYVDNRNRQOTLQGITPYQTDTY-----AQANTSTFALDLSLTLRA 481
Qy 437 TLGNYRNEYGNK-----RPEELGLFFDGDQDNGLYSYLGRFGKDGKLLPQ 485
Db 482 NYGLEFFYDKVRPSSQPRASTSAVGFPAAAGM---TPKGRALGSLFARLDYD----- 532
Qy 486 KSTIVQAGSOFNTFYDAALKDKIYRL-----NYSTNTVGYRFGGEYTCYGSDDDEF 539
Db 533 -----YDDLNLNLAGLRYDRYRLGDTGFNARTFELGTTRQDMPLOYAVD--- 578
Qy 540 KRAFGSPYTKKCNQSCGIYEPVLKKYKKRANNSHVSISADFG--DYFMPFASYSRTH 598
Db 579 -REEGRFSPTF-----GLSVKPGVDWLQLFATYKKGW 609
Qy 599 RMPNIOEMYSQIGDSG-----VHTALKPERANTWQFGENTYKGLLKQDDTLGLKLVG 652
Db 610 RPPAVTESLITGRPHGGAENWPNPFLSPERSKAWEGFNVLENLWFSDDRLGLKAVY 669
Qy 653 YRSRIDNYIHNVYKGMWDLNIPSWSSSTGLAYTIQHRNFKDKVHKHGFLELNLNDYGR 712
Db 670 FDRVDDDFIENGGMQ-----PPGYCMAGIGNSAYVNNL-DSTRFRGVEYQLDYDAGL 721
Qy 713 FFTNLISYAKS-----TQPTNPSDASESPNNASKEDQLKQYGLSRV--- 755
Db 722 AYQGLSYTHMIGNSDFCSKTAWLGVTQTVKSGRRPVPVDMRFPDQANAAHCSAVLGS 781
Qy 756 -SALPRDYGRLEVGRWLGNKLTLLGGAMRYFCGKSIRATAEERYIDGTNGGNTSNVRLGK 814
Db 782 AEHPMDRG-----SLTGL--NRFFDRDLVGCARARYSEGYSVAGCATVSQAGV 828
Qy 815 RSIKOTETLARQPLIFDYAAEPKKNLIFRAEVKNLDRRIYIDPLDAGNDAATQRYSS 874
Db 829 YPADWKEY-----TVYDLGYSYRVSDELTLRLAMENVTDRAVLVPLG----- 870
Qy 875 FDPDKDDEVTCNADKTLICNGKYGCTSKSVLT-NFARGTFLITMSYKF 922
Db 871 -----DVLAFTLGRGRTLOGTLEYQF 891

RESULT 9
Q9HUR6
ID Q9HUR6 PRELIMINARY; PRT: 989 AA.
AC Q9HUR6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HYPOTHEICAL PROTEIN PA4897.
GN PA4897.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004902; AAG08282.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 989 AA; 108303 MW; 99D33D2FBD0F0806F CRC64;
```

```
Query Match 8.7%; Score 425; DB 16; Length 989;
Best Local Similarity 23.2%; Pred. No. 1.8e-15;
Matches 230; Conservative 130; Mismatches 382; Indels 250; Gaps 42;

Qy 30 RAGSEAQIQLVEDVHVAKKRVKDKKVFETDARAVSTRODIFKSSNLD-----NIVR 81
Db 149 RMSGEAPADLSPVVVSAEALADPKETYT-----APRSSVYLSSEDIORFGRVSGDLLQ 203
Qy 82 SIPGAFTQODKSSGIVSLNIRGDSGFGRVNTWVGITOTFTYSTSDAGRAGSSQFASV 141
Db 204 GIPGVQVGDNRGALDINIRGICQSRVAVRVDAEQ-----ALDVRVAGTQOORSYI 258
Qy 142 DSNFTAGIDVYKVGSGSAGINSLAGSANLRTLGVDDVVOGNNYGLLLKGLTGTNSTKG 201
Db 259 DPDLVSSVTVDKGPSTRSGAI---GGSVEMRTIGVKDILVDGKDLGVRTGDVWNN--G 312
Qy 202 NAMAAGARKWLESGASV---GVLYGHSRR--SVAQNYRVGGGQHGNGFABYLRR 254
Db 313 VAPQHRASAKSTENLSSVPHDRGSLFGSAKSGSAFAFR---NEHL-DLVAAYAQRN 367
Qy 255 KORYFVQGGGLKFNNSGKWERDFQRPYK-----TKWY---QKYNPQELQKYE 302
Db 368 QGNF-----SOKKGODRYRVNRYGRESSVAKYNAGEEVLNSSSETESL- 415
Qy 303 GHDKSWRENLAPOYDIT-----PIDPSSLKQOAGNLFKLEYDGV-FNKYTAQF 350
Db 416 -LKATWR--IADEHTDLGYRRYDGRTEGIMPESDIFREGTAGIYOYPLSEVKIDITYARY 472
Qy 351 RDLNTKIGSRKIINRNYQFNGLSINNYANLNLTAAVNSGRQKYPKSGKFTWGLLKDFE 410
Db 473 RYLPEN-----NPLVDLSTGLWMTAKSDMLTSLVLAQRSQAYRSRDNWT----- 516
Qy 411 TYNNAKI-LDLNNTATFRL-----PRETELOTLGFN-----YF 443
Db 517 RQDNRTGGDLNNVARFETDFGDKLDLGGSFQVEDIOPKSVVYTLHDINARLRLDAT 576
Qy 444 HNEYGKN-----RPPEELGLFFDG-----PDQNGLYSYLGRFGKDKGLLPKQSTIVOPA 493
Db 577 ROEYGLNGKLEFKPVERLTLWGGYRSHFNKSDNGISAPREDRDMRFI-----TVSRPG 632
Qy 494 --GSQYF---NTFYFDAA---LKKDIYRLAYSTNTVGYREGGEYTYGYSDDDEKRAFQ 544
Db 633 YYGSMWMPDQNGYTDATDPRLNGIVTNTNPNFEGIPF-----DEFGPA-- 679
Qy 545 ENSPTYKKHKQSCGIYEPVLKKYKKRANNSHVSISADF---GDYFMPFASYSRTHRP 601
Db 680 -NVTVHPSRVNTVGTYN--YSKKGSSRGGGFSFPAGINFELAPDTFV-YASYTEGLRLP 735
Qy 602 NIQEMYFSQIGDSGVHITA--LKPERANTWQFGNTYKGLLKQDDTLGLKLVGYSRIDN 659
Db 736 SLFE---TSQGTQLQVPGKDLKPEERSWSWICASALRSLDADGSDAAIKLAYFNNTIKN 792
Qy 660 YIHNVYKGMWDLNNGNIPSWVSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFFTNLSY 719
Db 793 YITRYD-----PGQMLMTFSNT-----DSTRTSGLQLSHYDAGRVEFADUSA 836
Qy 720 AYQKSTQPTNFSADSESPNNASKEDQLKQYGLSRVSL-----PRDYGRELEVGR 770
Db 837 TYLLKTETCDAFAARLARAGANRYORTENTPNCITPSFGMSYNTQNPRLATNLTAGLR 896
Qy 771 WLGNKLTGGAMRYFGKKSIRATAEERYIDGTNGGNTSNVRLGKRSIKOTETLARQPLIF 830
Db 897 FFDQALTLGGRMTY--TSGPTATADKPMQVG-----ATTPOIYERSVQ-----LF 939
Qy 831 DFYAAEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFPDKDKDEVTCNADK 890
Db 940 DLFLKYLKLFELTNASLQNLTDRIYLDPL-----AQSPMP----- 975
Qy 891 TLCNGKYGCTSKSVLTNFARGTFLITMSYKF 922
Db 976 -----APGETLAVGMAQKF 989
```



```
QY 168 SANRLTGVDDVVGQNTTYGLLLKGLTGTNSTK--GNAMAAIGARKWLESGASGVLYGH 225
Db 259 IATFNTVSDSLAPGKELGCKLHASTGDNTHFIGSGILALGN---ETG---DILLAA 311
QY 226 SRSVAQNYRVGGGQHIGFAGAEYLERRRKQRYFVBOGGGLKNSNGSKWERDFQRPWKT 285
Db 312 SERHLG-DYWPGNKGD-IGN-----TRINDTGN----- 338
QY 286 KWKYQYNDOELOQYIEGHDK-----SWRENLAPOYDITPIDPSSLKOQS-----AGN 333
Db 339 --YDRVAESIKNNKIPDTHYRMSRLAKVGM--NLPAN---QRLQLSYLQTQASPAGT 391
QY 334 LFKLEYDVGNKNTAQPRDLNTKIGSRKIINRYQFNYGLSLANSYANLNTA-AYNSGRQ 392
Db 392 LTNL---GTRPPYELGWK---RTGYTDVNARNAAFDYSLAPEDVDWLDFQAKLYYVDTQ 444
QY 393 KYPKGSFTGWLGLKDPETYNNAKILDLNNTATFLPRE--TELQTLGFNY-----FHN 445
Db 445 -----DSDTYTSTSLD--NGYATRTLRTYCAQAQNTSRSLAPGHDFRA 489
QY 446 EYGNRPPEELGLFFDGPDDONGLYSYLGRFKDGKLLPKOKSTIVQAGSOYFNTFY--- 502
Db 490 NYG-----LEFYDKATSDSSR-----QGMGVTTPAGNRSV---ASLFANLTYDYD 532
QY 503 ----FDAALKDYLRLNYSNTVGY----REFGEYTYGYSDDDEK--RAFGENSPYKK 552
Db 533 GWLTFEGGLRYDRLRGQTG--LSYDPLAKDQORYT----IDNPKALRTLGCSTTTRED 587
QY 553 -HCNQSGIYEPVLKKYGKKRANHSHYSISADFGDFMPFASYSRTHRMFNIQEMVFSQI 611
Db 588 WDVDROGKLSPTL-----AVAVRPGV-EMLEYTYGKSWRPAITE---TL 631
QY 612 GDSGVHTA-----LKPERANTWQGFNTYKGLLKQDDTLGLKLVGYSRIDNVIHN 663
Db 632 TNGSAHSSSTQYPNPFQLQERSAWEGFNVQOPDLWFEGDRLVAKYGYDFTKVDNTINL 691
QY 664 VYKAWDLNGLNIPSWVS-SGLVATYIQHRNFKDKVHKHGFELNLYDYGREFTNLSVAYQ 722
Db 692 A-----IDRNKFGVPSIGNRAYV---NNLSKTRFEGLEYQLINDAGVYADLTTHM 742
QY 723 --KSTQPTN-----FSDASESPN-----NASKEDQLKQGYGLSRVSA--LPRDGR 764
Db 743 IGRNEFCSNKAWLGGRLRYGDGSRGNFYVEPDAAANDFTVCDGGTQFGSAAVLPGRDGS 802
QY 765 LEVGTRLHKLILGGAMRYFGKSIKRAETAEERYIDGTNGGNTSNVROLGKRSTKQETLA 824
Db 803 VTLGGRAFDRKLDAGVTVPF-----APGYQDSSVPSNYPYLADWPKYT----- 845
QY 825 RQPLIFDYAAIEPKNLIFRAEVKNLFDRLRYIDPLDAGNDAATQRYSSYFDPKDKDEEV 884
Db 846 ----LFDLYASYKLTDLSLTLGSGVENLTNRAYV----- 874
QY 885 TCNADKTLCKNGYGGTSKSVLTNPARGRTFLITMSYKF 922
Db 875 -----VSYGETLANTL---GRGRTVQGGVEYRF 899

RESULT 12
Q9CKJ4 ID Q9CKJ4 PRELIMINARY; PRT; 848 AA.
AC O9CKJ4;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE HASR.
GN HASR OR PM1622.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PW70;
```

```
RX MEDLINE-21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL: A5006199; AAK03706.1; -.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 848 AA; 95909 MW; 5C6B28E913F1D583 CRC64;

Query Match 7.2%; Score 351; DB 16; Length 848;
Best Local Similarity 21.7%; Pred. No. 1.9e-11;
Matches 207; Conservative 154; Mismatches 37; Indels 222; Gaps 48;

QY 12 FYLMGVMLYHSHVAEDAGRAGSEAOIOVEDVHVKAARVPKDK-----KVFEDARAVS- 64
Db 14 FLMLGLPTWAFSQAN-----LEKSTINKLETILVNESE-EKNFEDENLIKTYLSSGSYSY 67
QY 65 -TRQDI--FKSSENLDNIVRSIPGAFTQODKSSGIVSLNIRGDSGFGRVNTMWDGITQTF 121
Db 68 LSQSDISTFRGS-SVGDFLSGVPGVIVGNKRNSGALSVNIRGIANENRVPWID---KGL 123
QY 122 YSTSTDAGRAGSSQFCASVDSNFIAGLDVYVKFSFGSAGINSLAGSANLRTILGVDVVQ 181
Db 124 QSVPSYQYQAGSSTR--TYLDPDLSIQVEIEKGPSLQMDATGATGGVVRVETLRWQDIIP 181
QY 182 GNNTYIGLLLAGLTGNTS-----TKGNAMAAIGARKWLESGASGVLYGHS-----RRSV 230
Db 182 QGRNWGVRLALGTMTNIVSPPPYTRGYQT-----KYISK-----CLSNHTGLCQTY 231
QY 231 AQNVYVGGGGQHGHNFG-----AYLERRKQRYFVQEGGLKFNNSNSKWE 275
Db 232 APNARYSHSGFDLNAYNYSLAFANKWQADLVLAYAKRKGNYFVRHGQTPVIESIEFE 291
QY 276 RDQRPYWKTKWKYKNDPOELOKYIEGHDKSWRENLAPOYDITPIDPS-SLKQOSAGNL 334
Db 292 ED-----SVEVEKREPR-VHEEVEIGSLTFKENRSTLY--RPGEALNTSQDNTSYL 338
QY 335 FKLE-YDGV-----FNKYTAQFRDLNTKI-----GSRKIINRYQFNYGLS- 374
Db 339 AKINVYNDVHRLGLAYRHYHSRFEGEINSSILNFRAYGALQEGTEVKVD-SYHANYSNP 397
QY 375 LNSYANLNLAAYNSGRQKYPKGSKFT-----GWGLLKDPETY--NNAKILDNNATTF 426
Db 398 TTPVYVNLNVNAYETDS-----DSSNFTPFIEEYGYSLSSRHAHFLVSKQKGLSIENTSIF 452
QY 427 RL-PRETELOTTLGFNFHNEYGNR-----FPEELGLFFDGPDDQDNGLYSYLGRFK 477
Db 453 QLNDKPSLYKYLGAHSERYIQPRNAQARVRAKYPEDA---IGP-----LYTRDKRK 503
QY 478 GDKGLL-----POKSTIVQAGSOYFNTFYDAAALKDKIYRLNYSNTVGYRFGGEYTYG 533
Db 504 EWSAFVAANYPIITSLWKADIGLRYLQSTIYD-----YIVRTERVNIIGAFVNP 552
QY 534 GSD-----DEFKRAFGENSPYKHKCNQSGI-YEPVLKKYGKKRANHSHYSISADFGDY 587
Db 553 NPGNIWVEYKIDVYHKQAPVKNKGMSPVIMLTPEPI-----NGVQI----- 594
QY 588 FMPFASYSRTHRMFNIQEMVFSQIGDSGVHTA-----LKPERANTWQGFNTYKGLLKQ 642
Db 595 ---YTKYAEALRSPS---LFQATKGWSMSATADNLEQLRPERAKNWNWAGINLNFYENLGK 648
QY 643 DDTLGLKLVGYSRIDNYIHNVYKQWDLNGLNIPSWYSSTGLAYTIOHRNFKDKVHKHGF 702
Db 649 DNILGFLAYFNKKIKDILTRTSP-----KDKVTQT---INIQSAQFK-----GI 691
QY 703 EELNLYDYGREFTNLSYAYOKSTOPTNFSASESPNNASKEDQLKQGY----GLSRYSAL 758
Db 692 ELSAYIDMGKFAKLAGCTYTKTK---FCLTAE---QAGRGEQCNCSGYIYRSNLNN-AVP 744
QY 759 PRDYGRLEVGTWRLGNKLTGGMARYFGKS--IRATAERYIDGTNGGNTSNVROLGKRSI 817
```



```
Db 745 PRLNLHATGLTRLEFQKLDICARYSYSKRLVPVLSAERFV-----NTSSIE----- 791
QY 818 KQETTLARQPLIFDYAYAEKPKNLI FRAEVKNLFDRRYIDPLDAGNDAATQR 870
Db 792 -----WAPYSLV-DLYANYNVSNLKLTMMDNVFNRYIIDNNMGLNTAPGR 838

RESULT 13
Q98L70
ID Q98L70 PRELIMINARY; PRT; 747 AA.
AC Q98L70;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE HEME ACQUISITION PROTEIN, HASR.
GN MUR1155.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RC MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002996; BAB48593.1; -.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 747 AA; 80102 MW; 4105DA001B73E752 CRC64;

Query Match 6.9%; Score 339.5; DB 16; Length 747;
Best Local Similarity 20.0%; Pred. No. 6.9e-11;
Matches 189; Conservative 131; Mismatches 282; Indels 341; Gaps 38;

QY 23 SYAEDAGRAGEAQIQVLEDVHVHAKRVPKDK-----KVFTDARAVSTRQDIFKS 72
Db 52 AHAAQATQPAGEQTDQ-----SKKADQEKATAGATLLDKILVISRTGETAIESLAS 103
QY 73 SENLD-----NIVRSIPGAFTQODKSSGGIVSLNIRGDSGFGGRVNTWDGKITQ 119
Db 104 ASHVDQEQLAHRMATTPTNEMLLGPGVATQADARRVSTINSIRGLQDFGRVAVIVDQARQ 163
QY 120 TFYSTSTDAGRAGSSQFAGASVDSNFTAGDVKVSGSGSAGINSLAGSANLRTLGVDDV 179
Db 164 DFQ-----RSDHGTOSTYIDPELVKSVDVIRGVPANTYGSAGGVVFFDTKDAADF 216
QY 180 VQGNNTYGLLLKGLTGNSTKGNAMAIGAARKWLESAGYGV-VLYGHRSRRSAQNTRVGG 238
Db 217 LKPEETWASVTGRYESNG-KGWTSATGAYRENNWDALGNIVYRN-----YDNYK-DG 269
QY 239 GGQHIGNFAGAEYLERRKQRFVQEGGLKFNNSCKWEDRQRPYKWKYQKNDQBELQ 298
Db 270 GGDVNTGTGFDVL-----SGLLKTSI-----RP-----TENSELK 299
QY 299 KYIEGHDKSNRENLAPOYDITPDPSLKKQSQAGNLFKLEYDGVFNKYTAQFRDLNLTIG 358
Db 300 LGWVGSSDGHDETSGGM-----PVNDVDLKS-----NTFTA----- 330
QY 359 SRKIINRYQFNGLSUNYANLNLTAAYNYSGRQKYPKSGKFTGWGLLKDFETYNNAKIL 418
Db 331 -----RYNITDEKSLDLHLINTSYN----- 351
QY 419 DLNN*ATFRLEPRETELOTTIGFNFHNEYKGNRRPELGLFFDGPQDQDGLNGLYLGFRFKG 478
```

```
Db 352 -----KTNLDLT----- 358
QY 479 DKGLLPQKS-----TIVQAGSOYFNTFYFDAALKK-DIYRLN-YSTNTVGYR--FGG 527
Db 359 --SLVPQKREDPTITGLPTVLPAGSQ-----STFDVGTGTIDINWTSRPFETGGIAHETLYGG 412
QY 528 EYTGYYGSDDEFK---RAFGEN--SPTYKKKHQSCGIYE-----PVLKKYKGRANNNH 576
Db 413 DWVG-----DDVKTGGTAGGDSFYTPSGKRNVS---GAYVQDKLTWDNLEVIAGLRDNY 464
QY 577 SV--SISADFGDYMP-----FASYSRTHRMNPNIQEMYSFGIDSG 615
Db 465 SLKDSRTSGDRLSPRITVGVSPFESAGLAGLQFYGTVAEGYRSPSLTETLISGNHPAG 524
QY 616 V-----HTALKPERANTWOGFNTYKGLLKQDDTLGLKLVGYRSRIDNYIHNVYKWKW 669
Db 525 VTFPFLPNPNLRPETGKTTEFGINRQNDIFEPGDALRVAAAYFHNNDVDDYIDGVTLSPF 584
QY 670 DLNNGNTPSWSSSTGLAVTIQHRNFKDKVHKHGFELNLDYDGRFFTNLSYAYQKSTQPTN 729
Db 585 APGSGCP---FGGIPICFYQNF-AQKIDGFELEGVYDAGWGYAGLSAS----- 631
QY 730 FSDASEPNNAKEDQLKQGYLS-----RVSALPRDYGRLVGVTRWLGKLTLLGGAMR 783
Db 632 -----ITNGHTISYKGVADLATIPSSQVTAQGLRFLDKLTVGGEVQ 675
QY 784 YFGKSTRATAERYIIDGTNGGNTSNVRQLCKRSIKQTETLARQPLIEDFYAAAYEPKKNL- 842
Db 676 YNKGK-----KGN-----AVAEDYTLVNAFASYQATDNLK 705
QY 843 -IFRAEVKNLFDRRYIDPLDAGND-----AATQRY 871
Db 706 VDRAD--NLFDVKYANPLNGSTTVAYVEPGITLKAATMRF 745

RESULT 14
Q93SH4
ID Q93SH4 PRELIMINARY; PRT; 782 AA.
AC Q93SH4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HMUR PROTEIN.
GN HMUR.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=110SPC4;
RA Nienaber A., Hennecke H., Fischer H.M.;
RT "Discovery of a haem uptake system in the soil bacterium
RT Bradyrhizobium japonicum.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ311165; CAC38746.1; -.
SQ SEQUENCE 782 AA; 83713 MW; F8FC1EBFE568AA87 CRC64;

Query Match 6.4%; Score 312.5; DB 2; Length 782;
Best Local Similarity 20.3%; Pred. No. 2.3e-09;
Matches 187; Conservative 113; Mismatches 308; Indels 311; Gaps 37;

QY 30 RAGSEAOIQVLEDVHVHAKRVPKDKKVFETDARA---VSTRODIFKSSLENLDNIVRSIPG 85
Db 71 QAQSAAPVQTLDTITVAA---TKTREADALAPVSSISLDQIQGLQPNRLSDVFHVPVC 127
QY 86 AFTQODKSSGIVSLNIRGDSGFGVRVNTWVDGITQTFYSTSTDAGRAGSSQFAGSVDNSF 145
Db 128 VSFQERGGDDPATVINIRGLQDFGRVAVVVDGARQNYORT---GHNANGSFF---LDPEL 180
QY 146 IAGLDVVVKSFSGSAGINSLAGSANLRTLGVDDVVOGNNTYGLLLKGLTGTNKTGNAMA 205
```



```

Db 181 IGGVDDVRGTANTYGGGATGGLVFTKDKINDVLRGGERWGVLDLSGYSNNRRGLGSV 240
QY 206 AIGARKWLESGASVGLYHSHRSRAVQNYRVGGGQIGNFGAEBYLRERRKQRYFVQSGGL 265
Db 241 FGGVRATPDVDFGGAYV-----RTQGNK-DGNGTEIGTNGQ-----VEAGLM 284
QY 266 KFNNSCKWDERFORPWYKTKYQKYNDOELOKYIEGHKSKWRENAP-OYDITPID-- 322
Db 285 KLT-----VRP-----ALGHEVKFGAVPDQYQYDQIGQPNRG 315
QY 323 -----PSSLKQKQAGNLT-----FKLEYDGVFNKYTAQPRDLNTRKSGKII 363
Db 316 PTTTOALLALNRGSSVYASDAKNYSGVTWNYSLPSNLDFDWHMSVY-----GNTDN 368
QY 364 NRYQFNYGLSLNSYAN-----LNLTAAYNSGRQYKPKSGKFTGWLKLDKFETYNNAKILD 419
Db 369 DQTKYHYGTTPSAYCNGFGNNVCGVGDKR-----GVVL-----NTYG-----VD 410
QY 420 LNNIATFRLPRETELOTLTLGTFNFHNEYGKNRPEELGLFEDGPDQDNGLYSLGRFGKD 479
Db 411 ANNTRENV-----GDMRNALTWGVDAFQDD 436
QY 480 -----KGLLPQKSTIVQAG-----SQYENTVFDA--ALKKDIYRLNY--STN 519
Db 437 VIITDSRG-----NSNITTPSGIRTVSGFLQKQNYST-WLEAVSAIRYDRYDLDSGKTS 491
QY 520 TVGYRFGGTYGYSDDEPKAFGENSPYKKHCNOSCGIYEPVLKYYKKKRRANNHVS 579
Db 492 TGGDRFSPKIT-----LGV-TPV----- 508
QY 580 ISADFGDYFEPFASRTHRWPNIOEMFYSGIDSGVHTA----- 619
Db 509 -----PGFPYYSIAGYRAPSTETVI-----SGAHATGGGPAFPVCPDGTAGLFCFL 557
QY 620 -----LKPERANTWFGENTYKKGILLKODDTLGLKLGVYRSRIDNYIHNVYKGMWDLNGNI 675
Db 558 PNPNLREVEGKNKEVGLNYDNIFSANDSFRGKINLFRNDVSDYI-----DIVASA 609
QY 676 PSWVSSTG-LAYTIQHRNFKDKVHKHGFLELNYDYGRFTNLSYAY-QKSTOPTNFSDA 733
Db 610 PVAVPPGSGFSQFYQYQNIAN-ARIRGFEAETMYDAGDFIGVAGHYIQGNVATNIGLA 668
QY 734 SESPNNAKEDQLKQVGLSRVSALPRDYGRLVGTWGLNKLTLGAMRYFGKSIKIRATA 793
Db 669 TITP-----RKVVTTGGVLLDRTLLTA 692
QY 794 EERYIDGTNGNTSNVQLGKRISIKQETILARQPLIDFYAAVEPKKNLIFRAEVKNLFD 853
Db 693 QWASYGPNNDVPAGYLPATGYE-----LVNLYITYNATRDIVLSASIDNLLN 739
QY 854 ---RRYIDPLDAGNDAATQ 869
Db 740 QYVRPYAIP-GSSTDGTQT 757

```

RESULT 15

```

Q9KLM6 ID Q9KLM6 PRELIMINARY; PRT; 784 AA.
AC Q9KLM6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TONB RECEPTOR-RELATED PROTEIN.
GN VCA0625.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

```

```

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004392; AAF96526.1; -.
DR TIGR; VCA0625; -.
DR InterPro; IPR000531; TonB_boxC.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 784 AA; 87975 MW; 605DCFF12B0CBB10 CRC64;

Query Match          6.3%; Score 308; DB 16; Length 784;
Best Local Similarity 21.2%; Pred.No. 41e-09;
Matches 202; Conservative 115; Mismatches 314; Indels 320; Gaps 46;

QY 6 RLKPICFYLMGVMLYHSHYAEADAGRAGSEAIQVLEDVHVKAKRVPKDKKVTTDARAVST 65
Db 36 QLKPL-FTLLPVVLSVVQAQE---NTEQAVDETIVH-----GQSILTDQR---T 79
QY 66 QDIFKSSENLDNIVRSIPGAFTOODKSSGIVS-----LNRIGDSGFGRVN 111
Db 80 RSLDLK-----VRGIANA---DIFSGITSVQSNMHNHEAGALDIGIRGVGEGRVP 127
QY 112 TWDVGTITFTVSTDAGRAGSSQFASVDSNFIAGLDVVVKGS-----FSGSAGINS 164
Db 128 IFIDGSLQ---SHTSRGYQGVSDR--TYIDTLLSLTVNKGATIESSPYASGAVG--- 179
QY 165 LAGSANLRTIGVDVVOGNNNTYGLLLKGLTGNTSTKGNAMAAIGARKWLESGASGVLYG 224
Db 180 --GVNATTLGIKDIKDDQAFGVVLKARANHNRTPDV-----SGDYSEOGQ-----YA 227
QY 225 HSRSSVAQNVYRG---GGGQHIGNFGA--EYLERRQRYFVQEGGLKFNSNSGKWEDF 278
Db 228 LDERGEHSATKHGSLMLGLGYQAESFNTVLAYSKRSGKNHF-----AGK----- 271
QY 279 QRPYWKTKYQKYNP-----QELQKYIEGHKSKWRENAPQYDITPIDPSSLKQKSAGN 333
Db 272 -----KGVEEQEPVVGQGVVNTSFESD-SW----- 298
QY 334 LFKLEYDGVFNKYTAQPRDLNTRKIRIINRNYQFNYGLSLNSYANLNTAAYNSGRQK 393
Db 299 LFKLASD---TGTAHNADFNRYRHAQKAGEVLMAYWYKSSDWEGN-----PYPDGKDR 349
QY 394 YPKGSKFTGWLKDKFETYNNAKILDLNNTATFRLPRETELOTLTLGTFNFH-----EY 447
Db 350 MPQ-----WGL-----GTAKV-----NTYSANYYPQDHPWLN 378
QY 448 GKNRFPPEELGLFDGPDQDNGLYSLGRFGKGLKGLLPQKSTIVQAPAGSQFNTFYFAA- 506
Db 379 NANFWYTEADL-----AQYGLWA-LG-----TNAEQFPAHYHNDRSG 415
QY 507 -----LKKDIYRLNYSNTVGYRFGGTYGYGSDDEKRAFGENSPYKKHCNQSC 558
Db 416 LSLTNETLLTQWPVRLNVLGAAQNERLSPEBDG---QTRFTKT----- 455
QY 559 GIYEPVLKYYKKKRRANNHVSISADFG-----DY----- 587
Db 456 ---VTSRHGKRTAQNLFANADIDYSPRLVQLGLNLHNAKSTDYQTKQQLDYKEKLDLL 510
QY 588 -----FMP-----FASYSRTHRMENIOEMVFS-OIGDSGVHTALKPERANTWQEGENTYK 636
Db 511 SEFTYALTPTQLFLKSSRYRMESLLETTLNNEVFSVNPYNPKPEQAWNNEVGVOFMA 570
QY 637 KGLLKQDDTLGLKLGVYRSRIDNYIHNVYKGMWDLNGNIPSWSVSSTGLAYTIQHRNFKDK 696
Db 571 SNSVLQDRLNLNLSVYFNRNSIKDFISG--GLAKTPG-MSEW--QANFTFT---NY-DK 620

```

